



Edition: 2018-02-19

## List of Inria's Research Teams

1. Project-Team ASAP
2. Project-Team ASCOLA
3. Team ASPI
4. Project-Team CAIRN
5. Project-Team CELTIQUE
6. Project-Team CIDRE
7. Project-Team DIONYSOS
8. Project-Team DIVERSE
9. Project-Team DYLISS
10. Project-Team FLUMINANCE
11. Project-Team GENSCALE
12. Project-Team HYBRID
13. Project-Team HYCOMES 406
14. Project-Team I4S
15. Project-Team IPSO
16. Project-Team KERDATA
17. Project-Team LACODAM
18. Project-Team LAGADIC
19. Project-Team LINKMEDIA
20. Project-Team MIMETIC
21. Project-Team MYRIADS
22. Project-Team PACAP
23. Project-Team PANAMA
24. Project-Team SERPICO
25. Project-Team SIROCCO
26. Project-Team SUMO
27. Team TACOMA
28. Team TAMIS
29. Project-Team TEA
30. Project-Team VISAGES

# **Project-Team ASAP**

## As Scalable As Possible: foundations of large scale dynamic distributed systems

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: Institut national des sciences appliquées de Rennes

**Université Rennes 1** 

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Distributed Systems and middleware

### **Table of contents**

1.	Personnel	<b>7</b>
2.	Overall Objectives	
	2.1. Decentralized personalization	8
	2.2. Scalability: Cloud computing meets p2p	8
	2.3. Privacy-aware decentralized computations	9
	2.4. Information dissemination over social networks	9
	2.5. Computability and efficiency of distributed systems	10
3.	Research Program	
	3.1. Theory of distributed systems	11
	3.2. Peer-to-peer overlay networks	11
	3.3. Epidemic protocols	11
	3.4. Malicious process behaviors	12
	3.5. Online social networks and recommender systems	12
4.	. Highlights of the Year	
5.	. New Software and Platforms	
	5.1. WebGC	12
	5.2. Asapknn (MediEgo)	13
	5.3. YALPS	13
	5.4. GossipLib	13
6.	New Results	
	6.1. Theory of Distributed Systems	14
	6.1.1. Simulation of Partial Replication in Distributed Transaction	nal Memory 14
	6.1.2. Distributed Universal Constructions: a Guided Tour	14
	6.1.3. Atomic Read/Write Memory in Signature-Free Byzantine A	synchronous Message-Passing
	Systems	14
	6.1.4. From wait-free to arbitrary concurrent solo executions in co	olorless distributed computing 15
	6.1.5. Early Decision and Stopping in Synchronous Consensus: A	Predicate-Based Guided Tour
		15
	6.1.6. Long-Lived Tasks	15
	6.1.7. Which Broadcast Abstraction Captures k-Set Agreement?	16
	6.1.8. Signature-free asynchronous Byzantine systems: from m	ultivalued to binary consensus
	with $t < n/3$ , O(n2) messages, and constant time.	16
	6.1.9. A distributed leader election algorithm in crash-recovery and	nd omissive system 16
	6.1.10. Providing Collision-Free and Conflict-Free Communica	tion in General Synchronous
	Broadcast/Receive Networks	16
	6.1.11. Randomized abortable mutual exclusion with constant amo	rtized RMR complexity on the
	CC model.	17
	6.2. Network and Graph Algorithms	17
	6.2.1. Tight bounds on vertex connectivity under sampling	17
	6.2.2. Tight bounds for coalescing-branching random walks on re	gular graphs 17
	6.3. Scalable Systems	18
	6.3.1. Agar: A Caching System for Erasure-Coded Data	18
	6.3.2. Filament: A Cohort Construction Service for Decentraliz	ed Collaborative Editing Plat-
	forms	18
	6.3.3. Scalable Anti-KNN: Decentralized Computation of k-F	urthest-Neighbor Graphs with
	HyFN	18
	6.3.4. Density and Mobility-driven Evaluation of Broadcast Algo	rithms for MANETs 19
	6.3.5. An Adaptive Peer-Sampling Protocol for Building Network	ts of Browsers 19
	6.3.6. Designing Overlay Networks for Decentralized Clouds	19

7.	Partners	hips and Cooperations	
	7.1. Na	tional Initiatives	20
	7.1.1.	ANR project SocioPlug	20
	7.1.2.	DeSceNt CominLabs	20
	7.1.3.	ANR Blanc project Displexity	20
	7.1.4.	ANR project PAMELA	21
	7.1.5.	ANR project OBrowser	21
	7.1.6.	ANR project DESCARTES	21
	7.1.7.	ANR-ERC Tremplin project NDFUSION	21
	7.2. Int	ernational Initiatives	22
	7.2.1.	Inria International Labs	22
	7.2.2.	Inria Associate Teams Not Involved in an Inria International Labs	22
	7.3. Int	ernational Research Visitors	22
	7.3.1.	Visits of International Scientists	22
	7.3.2.	Visits to International Teams	22
8.	Dissemin	ation	
	8.1. Pro	omoting Scientific Activities	23
	8.1.1.	Scientific Events Organisation	23
	8.1.2.	Scientific Events Selection	23
	8.1.3.	Journal	23
	8.1	.3.1. Member of the Editorial Boards	23
	8.1	.3.2. Reviewer - Reviewing Activities	23
	8.1.4.	Invited Talks	23
	8.1.5.	Leadership within the Scientific Community	24
	8.1.6.	Research Administration	24
	8.2. Te	aching - Supervision - Juries	24
	8.2.1.	Teaching	24
	8.2.2.	Supervision	25
	8.2.3.	Juries	25
	8.3. Po	pularization	25
9.	Bibliogra	phy	

#### **Project-Team ASAP**

*Creation of the Project-Team: 2007 July 01, end of the Project-Team: 2017 December 31* **Keywords:** 

#### **Computer Science and Digital Science:**

- A1.1.1. Multicore, Manycore
- A1.1.6. Cloud
- A1.1.7. Peer to peer
- A1.1.9. Fault tolerant systems
- A1.2.9. Social Networks
- A1.3. Distributed Systems
- A1.5.2. Communicating systems
- A2.1.6. Concurrent programming
- A2.1.7. Distributed programming
- A2.6.2. Middleware
- A3.1.3. Distributed data
- A3.1.8. Big data (production, storage, transfer)
- A3.5.1. Analysis of large graphs
- A3.5.2. Recommendation systems
- A4.8. Privacy-enhancing technologies
- A7.1. Algorithms
- A8.7. Graph theory

#### **Other Research Topics and Application Domains:**

- B6.3.1. Web
- B6.3.3. Network Management
- B6.3.4. Social Networks
- B6.4. Internet of things
- B6.5. Information systems
- B9.4.1. Computer science
- B9.8. Privacy

### 1. Personnel

#### **Research Scientists**

Davide Frey [Inria, Researcher] George Giakkoupis [Inria, Researcher] Anne-Marie Kermarrec [Inria, Senior Researcher, until Jan 2017, HDR]

#### **Faculty Members**

Francois Taiani [Team leader, Univ de Rennes I, Professor, HDR] Marin Bertier [INSA Rennes, Associate Professor] David Bromberg [Univ de Rennes I, Professor, HDR] Michel Raynal [Univ de Rennes I, Professor, HDR]

#### **PhD Students**

Simon Bouget [Univ de Rennes I, until Oct 2017] Louison Gitzinger [Univ de Rennes I, from Oct 2017] Adrien Luxey [Univ de Rennes I] Pierre-Louis Roman [Univ de Rennes I, until Nov 2017] Olivier Ruas [Inria]

#### **Post-Doctoral Fellow**

Zohir Bouzid [Univ de Rennes I, until May 2017]

Visiting Scientist Arvid Bosk [KTH Royal Institute of Technology]

Administrative Assistant Virginie Desroches [Inria]

## 2. Overall Objectives

#### 2.1. Decentralized personalization

Our first objective is to offer full-fledged personalization in notification systems. Today, almost everyone is suffering from an overload of information that hurts both users and content providers. This suggests that not only will notification systems take a prominent role but also that, in order to be useful, they should be personalized to each and every user depending on her activity, operations, posts, interests, etc. In the GOSSPLE implicit instant item recommender, through a simple interface, users get automatically notified of items of interest for them, without explicitly subscribing to feeds or interests. They simply have to let the system know whether they like the items they receive (typically through a like/dislike button). Throughout the system's operation the personal data of users is stored on their own machines, which makes it possible to provide a wide spectrum of privacy guarantees while enabling cross-application benefits.

Our goal here is to provide a fully decentralized solution without ever requiring users to reveal their private preferences.

#### 2.2. Scalability: Cloud computing meets p2p

Our second objective is to move forward in the area of **scalable infrastructures** for data intensive applications. In this context, we focus significant efforts on personalization systems, which represent one of the biggest challenges addressed by most large stake holders.

**Hybrid infrastructures for personalisation.** So far, social filtering techniques have mainly been implemented on centralized architectures relying on smart heuristics to cope with an increasing load of information. We argue however that, no matter how smart these heuristics and how powerful the underlying machines running them, a fully centralized approach might not be able to cope with the exponential growth of the Internet and, even if it does, the price to be paid might simply not be acceptable for its users (privacy, ecological footprint, etc.).

At the other end of the spectrum, lie fully decentralized systems where the collaborative filtering system is implemented by the machines of the users themselves. Such approaches are appealing for both scalability and privacy reasons. With respect to scalability, storage and computational units naturally grow with the number of users. Furthermore, a p2p system provides an energy-friendly environment where every user can feel responsible for the ecological foot-print of her exploration of the Internet. With respect to privacy, users are responsible for the management of their own profiles. Potential privacy threats therefore do not come from a big-brother but may still arise due to the presence of other users.

We have a strong experience in devising and experimenting with such kinds of p2p systems for various forms of personalization. More specifically, we have shown that personalization can be effective while maintaining a reasonable level of privacy. Nevertheless, frequent connections/disconnections of users make such systems difficult to maintain while addressing privacy attacks. For this reason, we also plan to explore hybrid approaches where the social filtering is performed by the users themselves, as in a p2p manner, whereas the management of connections-disconnections, including authentication, is managed through a server-based architecture. In particular, we plan to explore the trade-off between the quality of the personalization process, its efficiency and the privacy guarantees.

#### 2.3. Privacy-aware decentralized computations

Gossip algorithms have also been studied for more complex global tasks, such as computation of network statistics or, more generally, aggregation functions of input values of the nodes (e.g., sum, average, or max). We plan to pursue this research direction both from a theoretical and from a practical perspective. We provide two examples of these directions below.

**Computational capabilities of gossip**. On the theoretical side, we have recently started to study gossip protocols for the assignment of unique IDs from a small range to all nodes (known as the *renaming* problem) and computing the rank of the input value of each node. We plan to further investigate the class of global tasks that can be solved efficiently by gossip protocols.

**Private computations on decentralized data.** On a more practical track, we aim to explore the use of gossip protocols for decentralized computations on privacy sensitive data. Recent research on private data bases, and on homomorphic encryption, has demonstrated the possibility to perform complex operations on encrypted data. Yet, existing systems have concentrated on relatively small-scale applications. In the coming years, we instead plan to investigate the possibility to build a framework for querying and performing operations for large-scale decentralized data stores. To achieve this, we plan to disseminate queries in an epidemic fashion through a network of data sources distributed on a large scale while combining privacy preserving techniques with decentralized computations. This would, for example, enable the computation of statistical measures on large quantities of data without needing to access and disclose each single data item.

#### 2.4. Information dissemination over social networks

While we have been studying information dissemination in practical settings (such as WhatsUp in GOSSPLE), modeling such dynamic systems is still in its infancy. We plan to complement our practical work on gossip algorithms and information dissemination along the following axes:

**Rumour spreading** is a family of simple randomized algorithms for information dissemination, in which nodes contact (uniformly) random neighbours to exchange information with them. Despite their simplicity these protocols have proved very efficient for various network topologies. We are interested in studying their properties in specific topologies such as social networks be they implicit (interest-based as in GOSSPLE) or explicit (where users choose their friends as in Facebook). Recently, there has been some work on bounding the speed of rumour spreading in terms of abstract properties of the network graph, especially the graph's expansion properties of conductance and vertex expansion. It has been shown that high values for either of these guarantees fast rumour spreading—this should be related to empirical observations that social networks have high expansion. Some works established increasingly tighter upper bounds for rumour spreading in term of conductance or vertex expansion, but these bounds are not tight.

Our objective is to prove the missing tight upper bound for rumour spreading with vertex expansion. It is known that neither conductance nor vertex expansion are enough by themselves to completely characterize the speed of rumour spreading: are there graphs with bad expansion in which rumours spread fast?

**Overcoming the dependence on expansion**: Rumour spreading algorithms have very nice properties such as their simplicity, good performances for many networks but they may have very poor performance for some networks, even though these networks have small diameter, and thus it is possible to achieve fast information dissemination with more sophisticated protocols. Typically nodes may choose the neighbours to contact with some non-uniform probabilities that are determined based on information accumulated by each node during the run of the algorithm. These algorithms achieve information dissemination in time that is close to the diameter of the network. These algorithms, however, do not meet some of the other nice properties of rumour spreading, most importantly, robustness against failures. We are investigating algorithms that combine the good runtime of these latest protocols with the robustness of rumour spreading.

**Competing rumours**: Suppose now that two, or more, conflicting rumours (or opinions) spread in the network, and whenever a node receives different rumours it keeps only one of them. Which rumour prevails, and how long does it take until this happens? Similar questions have been studied in other contexts but not in the context of rumour spreading. The *voter* model is a well studied graph process that can be viewed as a competing rumour process that follows the classic PULL rumour spreading algorithm. However, research has only recently started to address the question of how long it takes until a rumour prevails. An interesting variant of the problem that has not been considered before is when different rumours are associated with different weights (some rumours are more convincing than others). We plan to study the above models and variations of them, and investigate their connection to the standard rumour spreading algorithms. This is clearly related to the dissemination of news and personalization in social networks.

#### **2.5.** Computability and efficiency of distributed systems

A very relevant challenge (maybe a Holy Grail) lies in the definition of a computation model appropriate to dynamic systems. This is a fundamental question. As an example there are a lot of peer-to-peer protocols but none of them is formally defined with respect to an underlying computing model. Similarly to the work of Lamport on "static" systems, a model has to be defined for dynamic systems. This theoretical research is a necessary condition if one wants to understand the behavior of these systems. As the aim of a theory is to codify knowledge in order it can be transmitted, the definition of a realistic model for dynamic systems is inescapable whatever the aim we have in mind, be it teaching, research or engineering.

**Distributed computability**: Among the fundamental theoretical results of distributed computing, there is a list of problems (e.g., consensus or non-blocking atomic commit) that have been proved to have no deterministic solution in asynchronous distributed computing systems prone to failures. In order such a problem to become solvable in an asynchronous distributed system, that system has to be enriched with an appropriate oracle (also called failure detector). We have been deeply involved in this research and designed optimal consensus algorithms suited to different kind of oracles. This line of research paves the way to rank the distributed computing problems according to the "power" of the additional oracle they required (think of "additional oracle" as "additional assumptions"). The ultimate goal would be the statement of a distributed computing hierarchy, according to the minimal assumptions needed to solve distributed computing problems (similarly to the Chomsky's hierarchy that ranks problems/languages according to the type of automaton they need to be solved).

**Distributed computing abstractions**: Major advances in sequential computing came from machineindependent data abstractions such as sets, records, etc., control abstractions such as while, if, etc., and modular constructs such as functions and procedures. Today, we can no longer envisage not to use these abstractions. In the "static" distributed computing field, some abstractions have been promoted and proved to be useful. Reliable broadcast, consensus, interactive consistency are some examples of such abstractions. These abstractions have well-defined specifications. There are both a lot of theoretical results on them (mainly decidability and lower bounds), and numerous implementations. There is no such equivalent for dynamic distributed systems, i.e. for systems characterized by nodes that may join and leave, or that may change their characteristics at runtime. Our goal is to define such novel abstractions, thereby extending the theory of distributed systems to the dynamic case.

## 3. Research Program

#### **3.1.** Theory of distributed systems

Finding models for distributed computations prone to asynchrony and failures has received a lot of attention. A lot of research in this domain focuses on what can be computed in such models, and, when a problem can be solved, what are its best solutions in terms of relevant cost criteria. An important part of that research is focused on distributed computability: what can be computed when failure detectors are combined with conditions on process input values for example. Another part is devoted to model equivalence. What can be computed with a given class of failure detectors? Which synchronization primitives is a given failure class equivalent to? These are among the main topics addressed in the leading distributed computing community. A second fundamental issue related to distributed models is the definition of appropriate models suited to dynamic systems. Up to now, the researchers in that area consider that nodes can enter and leave the system, but do not provide a simple characterization, based on properties of computation instead of description of possible behaviors [46], [40], [41]. This shows that finding dynamic distributed computing models is today a "Holy Grail", whose discovery would allow a better understanding of the essential nature of dynamic systems.

#### 3.2. Peer-to-peer overlay networks

A standard distributed system today is related to thousands or even millions of computing entities scattered all over the world and dealing with a huge amount of data. This major shift in scalability requirements has lead to the emergence of novel computing paradigms. In particular, the peer-to-peer communication paradigm imposed itself as the prevalent model to cope with the requirements of large scale distributed systems. Peer-topeer systems rely on a symmetric communication model where peers are potentially both clients and servers. They are fully decentralized, thus avoiding the bottleneck imposed by the presence of servers in traditional systems. They are highly resilient to peers arrivals and departures. Finally, individual peer behavior is based on a local knowledge of the system and yet the system converges toward global properties.

A peer-to-peer overlay network logically connects peers on top of IP. Two main classes of such overlays dominate, structured and unstructured. The differences relate to the choice of the neighbors in the overlay, and the presence of an underlying naming structure. Overlay networks represent the main approach to build large-scale distributed systems that we retained. An overlay network forms a logical structure connecting participating entities on top of the physical network, be it IP or a wireless network. Such an overlay might form a structured overlay network [47], [48], [49] following a specific topology or an unstructured network [45], [50] where participating entities are connected in a random or pseudo-random fashion. In between, lie weakly structured peer-to-peer overlays where nodes are linked depending on a proximity measure providing more flexibility than structured overlays and better performance than fully unstructured ones. Proximity-aware overlays connect participating entities so that they are connected to close neighbors according to a given proximity metric reflecting some degree of affinity (computation, interest, etc.) between peers. We extensively use this approach to provide algorithmic foundations of large-scale dynamic systems.

#### **3.3. Epidemic protocols**

Epidemic algorithms, also called gossip-based algorithms [44], [43], constitute a fundamental topic in our research. In the context of distributed systems, epidemic protocols are mainly used to create overlay networks and to ensure a reliable information dissemination in a large-scale distributed system. The principle underlying technique, in analogy with the spread of a rumor among humans via gossiping, is that participating entities continuously exchange information about the system in order to spread it gradually and reliably. Epidemic algorithms have proved efficient to build and maintain large-scale distributed systems in the context of many applications such as broadcasting [43], monitoring, resource management, search, and more generally in building unstructured peer-to-peer networks.

#### **3.4.** Malicious process behaviors

When assuming that processes fail by simply crashing, bounds on resiliency (maximum number of processes that may crash, number of exchanged messages, number of communication steps, etc.) are known both for synchronous and augmented asynchronous systems (recall that in purely asynchronous systems some problems are impossible to solve). If processes can exhibit malicious behaviors, these bounds are seldom the same. Sometimes, it is even necessary to change the specification of the problem. For example, the consensus problem for correct processes does not make sense if some processes can exhibit a Byzantine behavior and thus propose an arbitrary value. In this case, the validity property of consensus, which is normally "a decided value is a proposed value", must be changed to "if all correct processes propose the same value then only this value can be decided." Moreover, the resilience bound of less than half of faulty processes is at least lowered to "less than a third of Byzantine processes." These are some of the aspects that underlie our studies in the context of the classical model of distributed systems, in peer-to-peer systems and in sensor networks.

#### 3.5. Online social networks and recommender systems

Social Networks have rapidly become a fundamental component of today's distributed applications. Web 2.0 applications have dramatically changed the way users interact with the Internet and with each other. The number of users of websites like Flickr, Delicious, Facebook, or MySpace is constantly growing, leading to significant technical challenges. On the one hand, these websites are called to handle enormous amounts of data. On the other hand, news continue to report the emergence of privacy threats to the personal data of social-network users. Our research aims to exploit our expertise in distributed systems to lead to a new generation of scalable, privacy-preserving, social applications.

We also investigate approaches to build implicit social networks, connecting users sharing similar interests. At the heart of the building of such similarity graphs lie k-nearest neighbor (KNN) algorithms. Our research in this area is to design and implement efficient KNN algorithms able to cope with a huge volume of data as well as a high level of dynamism. We investigate the use of such similarity graphs to build highly scalable infrastructures for recommendation systems.

## 4. Highlights of the Year

#### 4.1. Awards

• Anne-Marie Kermarrec received the Inria/Dassault Systems/Académie des science/ Innovation Award in 2017.

## 5. New Software and Platforms

#### 5.1. WebGC

#### Web-based Gossip Communication

KEYWORDS: WebRTC - Recommendation systems - Decentralized architectures - Personalized systems - Web - Peer-to-peer - Gossip protocols - Epidemic protocols - Decentralized web SCIENTIFIC DESCRIPTION: The library currently includes the implementation of two peer sampling protocols,

Cyclon and the generic peer-sampling protocol from, as well as a clustering protocol. All protocols implement a common GossipProtocol "interface" FUNCTIONAL DESCRIPTION: WebGC consists of a WebRTC-based library that supports gossip-based communication between web browsers and enables them to operate with Node-JS applications. WebGC comprises the implementation of standard gossip protocols such as Peer Sampling or Clustering, and simplifies the development of new protocols. It comprises a decentralized signaling service that makes it easier to build completely decentralized browser-based applications.

- Participants: Anne-Marie Kermarrec, Davide Frey, Matthieu Simonin and Raziel Carvajal Gomez
- Contact: Davide Frey

#### 5.2. Asapknn (MediEgo)

KEYWORDS: Widget web - Social network - Recommendation

FUNCTIONAL DESCRIPTION: Asapknn (MediEgo) is a solution for content recommendation based on the users navigation history. The solution 1) collects the usages of the Web users and store them in a profile, 2) uses this profile to associate to each user her most similar users, 3) leverages this implicit network of close users in order to infer their preferences and recommend advertisements and recommendations. MediEgo achieves scalability using a sampling method, which provides very good results at a drastically reduced cost.

- Participants: Anne Marie Kermarrec, Antoine Boutet, Arnaud Jegou, Davide Frey, Jacques Falcou, Jean-Francois Verdonck, Rachid Guerraoui and Sébastien Campion
- Partner: EPFL Ecole Polytechnique Fédérale de Lausanne
- Contact: Sébastien Campion

#### **5.3. YALPS**

KEYWORDS: Simulator - Peer-to-peer - Experimentation - Nat traversal - Traffic-shaping - Deployment FUNCTIONAL DESCRIPTION: YALPS is an open-source Java library designed to facilitate the development, deployment, and testing of distributed applications. Applications written using YALPS can be run both in simulation and in real-world mode without changing a line of code or even recompiling the sources. A simple change in a configuration file will load the application in the proper environment. A number of features make YALPS useful both for the design and evaluation of research prototypes and for the development of applications to be released to the public. Specifically, YALPS makes it possible to run the same application as a simulation or in a real deployment. Applications communicate by means of application-defined messages which are then routed either through UDP/TCP or through YALPS's simulation infrastructure. In both cases, YALPS's communication layer offers features for testing and evaluating distributed protocols and applications. Communication channels can be tuned to incorporate message losses or to constrain their outgoing bandwidth. Finally, YALPS includes facilities to support operation in the presence of NATs and firewalls using relaying and NAT-traversal techniques. The implementation of YALPS includes approximately 16K lines of code, and is used in several projects by ASAP, including HEAP, AllYours-P2P, and Behave.

- Participants: Anne Marie Kermarrec, Arnaud Jegou, Davide Frey, Heverson Borba Ribeiro and Maxime Monod
- Contact: Davide Frey
- URL: http://yalps.gforge.inria.fr/

#### 5.4. GossipLib

KEYWORDS: Nat traversal - Epidemic protocols - Gossip protocols - Overlay maintenance - Peer-to-peer - Dissemination

FUNCTIONAL DESCRIPTION: GossipLib is a library consisting of a set of Java classes aimed to facilitate the development of gossip-based application in a large-scale setting. It provides developers with a set of support classes that constitute a solid starting point for building any gossip-based application. GossipLib is designed to facilitate code reuse and testing of distributed application and as thus also provides the implementation of a number of standard gossip protocols that may be used out of the box or extended to build more complex protocols and applications. These include for example the peer-sampling protocols for overlay management.

GossipLib also provides facility for the configuration and deployment of applications as final-product but also as research prototype in environments like PlanetLab, clusters, network emulators, and even as event-based simulation. The code developed with GossipLib can be run both as a real application and in simulation simply by changing one line in a configuration file.

RELEASE FUNCTIONAL DESCRIPTION: Library for gossip-based applications and experiments

- Participants: Anne Marie Kermarrec, Davide Frey, Ilham Ikbal, Imane Al Ifdal and Ribeiro Heverson
- Contact: Davide Frey
- URL: http://gossiplib.gforge.inria.fr/

### 6. New Results

#### 6.1. Theory of Distributed Systems

#### 6.1.1. Simulation of Partial Replication in Distributed Transactional Memory Participant: François Taïani.

Distributed Transactional Memory (DTM) is a concurrency mechanism aimed at simplifying distributed programming by allowing operations to execute atomically, mirroring the well-known transaction model of relational databases. DTM can play a fundamental role in the coordination of participants in mobile distributed applications. Most DTM solutions follow a full replication scheme, in spite of recent studies showing that partial replication approaches can present gains in scalability by reducing the amount of data stored at each node. This work [33] investigates the role of replica location in DTMs. The goal is to understand the effect of latency on the DTM's system performance in face of judicious replica distribution, taking into consideration the locations where data is more frequently accessed.

This work was performed in collaboration with with Diogo Lima and Hugo Miranda from the University of Lisbon (Portugal).

#### 6.1.2. Distributed Universal Constructions: a Guided Tour

#### Participant: Michel Raynal.

The notion of a universal construction is central in computing science: the wheel has not to be reinvented for each new problem. In the context of n-process asynchronous distributed systems, a universal construction is an algorithm that is able to build any object defined by a sequential specification despite the occurrence of up to (n - 1) process crash failures. Michel Raynal presented a guided tour of such universal constructions in the bulletin of the EATCS [22]. Its spirit is not to be a catalog of the numerous constructions proposed so far, but a (as simple as possible) presentation of the basic concepts and mechanisms that constitute the basis these constructions rest on.

#### 6.1.3. Atomic Read/Write Memory in Signature-Free Byzantine Asynchronous Message-Passing Systems

Participant: Michel Raynal.

This work introduced a signature-free distributed algorithm which builds an atomic read/write shared memory on top of a fully connected peer-to-peer n-process asynchronous message-passing system in which up to t<n/3 processes may commit Byzantine failures. From a conceptual point of view, this algorithm is designed to be as close as possible to the algorithm proposed by [42], which builds an atomic register in an n-process asynchronous message-passing system where up to t<n/2 processes may crash. The proposed algorithm is particularly simple. It does not use cryptography to cope with Byzantine processes, and is optimal from a t-resilience point of view (t<n/3). A read operation requires O(n) messages, and a write operation requires  $O(n^2)$  messages. This work was done in collaboration with Achour Mostéfaoui, Matoula Petrolia and Claude Jard from the University of Nantes and was published in Theory of Computing Systems [19].

#### 6.1.4. From wait-free to arbitrary concurrent solo executions in colorless distributed computing Participant: Michel Raynal.

In an asynchronous distributed system where any number of processes may crash, a process may have to run solo, computing its local output without receiving any information from other processes. In the basic shared memory system where the processes communicate through atomic read/write registers, at most one process may run solo.

In this work we introduced a new family of d-solo models, where d-processes may concurrently run solo,  $1 \le d \le n$  (the 1-solo model is the basic read/write model). We studied distributed colorless computations in the d-solo models, where process ids are not used, either in task specifications or during computation, and we characterized the colorless tasks that can be solved in each d-solo model. Colorless tasks include consensus, set agreement and many other previously studied tasks. This shows that colorless algorithms have limited computational power for solving tasks, only when d>1. When d=1, colorless algorithms can solve the same tasks as algorithms that may use ids. It is well-known that, while consensus is not wait-free solvable in a model where at most one process may run solo,  $\epsilon$ -approximate agreement is solvable. In a d-solo model, the fundamental solvable task is  $(d,\epsilon)$ -solo approximate agreement, a generalization of  $\epsilon$ -approximate agreement. Indeed,  $(d,\epsilon)$ -solo approximate agreement can be solved in the d-solo model, but not in the (d+1)-solo model.

This work was carried out in collaboration with Maurice Herlihy from Brown University, Sergio Rajsbaum from UNAM (Mexico), and Julien Stainer from EPFL, in the context of the LIDICo associate team. It was published in Theoretical Computer Science [18].

#### 6.1.5. Early Decision and Stopping in Synchronous Consensus: A Predicate-Based Guided Tour

#### Participant: Michel Raynal.

Consensus is the most basic agreement problem encountered in fault-tolerant distributed computing: each process proposes a value and non-faulty processes must agree on the same value, which has to be one of the proposed values. While this problem is impossible to solve in asynchronous systems prone to process crash failures, it can be solved in synchronous (round-based) systems where all but one process might crash in any execution. It is well-known that (t + 1) rounds are necessary and sufficient in the worst case execution scenario for the processes to decide and stop executing, where t < n is a system parameter denoting the maximum number of allowed process crashes and n denotes the number of processes in the system. Early decision and stopping considers the case where f < t processes actually crash, f not being known by processes. It has been shown that the number of rounds that have to be executed in the worst case is then min(f + 2, t + 1). In this work we showed that this value is an upper bound attained only in worst execution scenarios. This work resulted from a collaboration with Armando Castaneda from UNAM, Yoram Moses from Technion, and Matthieu Roy from LAAS Toulouse, in the context of the LIDICo associate team. It was published at NETYS 2017 [29].

#### 6.1.6. Long-Lived Tasks

#### Participant: Michel Raynal.

The predominant notion for specifying problems to study distributed computability are tasks. Notable examples of tasks are consensus, set agreement, renaming and commit-adopt. The theory of task solvability is well-developed using topology techniques and distributed simulations. However, concurrent computing problems are usually specified by objects. Tasks and objects differ in at least two ways. While a task is a one-shot problem, an object, such as a queue or a stack, typically can be invoked multiple times by each process. Also, a task, defined in terms of sets, specifies its responses when invoked by each set of processes concurrently, while an object, defined in terms of sequences, specifies the outputs the object may produce when it is accessed sequentially.

In this work we showed how the notion of tasks can be extended to model any object. A potential benefit of this result is the use of topology, and other distributed computability techniques to study long-lived objects. This work resulted from a collaboration with Armando Castaneda and Sergio Rasjbaum from UNAM in the context of the LIDICo associate team. It was published at NETYS 2017 [35].

#### 6.1.7. Which Broadcast Abstraction Captures k-Set Agreement?

Participant: Michel Raynal.

It is well-known that consensus (one-set agreement) and total order broadcast are equivalent in asynchronous systems prone to process crash failures. Considering wait-free systems, we addressed and answered the following question: which is the communication abstraction that "captures" k-set agreement? To this end, we introduced a new broadcast communication abstraction, called k-BO-Broadcast, which restricts the disagreement on the local deliveries of the messages that have been broadcast (1-BO-Broadcast boils down to total order broadcast). Hence, in this context, k=1 is not a special number, but only the first integer in an increasing integer sequence. This establishes a new "correspondence" between distributed agreement problems and communication abstractions, which enriches our understanding of the relations linking fundamental issues of fault-tolerant distributed computing. This work was carried out in collaboration with Damien Imbs from the University of Marseille, Achour Mostéfaoui from the University of Nantes, and Matthieu Perrin from IMDEA (Spain). It was published at DISC 2017 [39].

# 6.1.8. Signature-free asynchronous Byzantine systems: from multivalued to binary consensus with t< n/3, O(n2) messages, and constant time.</li> Participant: Michel Raynal.

We introduced a new algorithm that reduces multivalued consensus to binary consensus in an asynchronous message-passing system made up of n processes where up to t may commit Byzantine failures. This algorithm has the following noteworthy properties: it assumes t < n/3t < n/3 (and is consequently optimal from a resilience point of view), uses  $O(n^2)$  messages, has a constant time complexity, and uses neither signatures nor additional computational power (such as random numbers, failure detectors, additional scheduling assumption, or additional synchrony assumption). The design of this reduction algorithm relies on two new all-to-all communication abstractions. The first one allows the non-faulty processes to reduce the number of proposed values to c, where c is a small constant. The second communication abstraction allows each non-faulty process to compute a set of (proposed) values satisfying the following property: if the set of a non-faulty process is a singleton containing value v, the set of any non-faulty process contains v. Both communication abstractions have an  $O(n^2)$  message complexity and a constant time complexity. The reduction of multivalued Byzantine consensus to binary Byzantine consensus is then a simple sequential use of these communication abstractions. To the best of our knowledge, this is the first asynchronous message-passing algorithm that reduces multivalued consensus to binary consensus with  $O(n^2)$  messages and constant time complexity (measured with the longest causal chain of messages) in the presence of up to t < n/3 t < n/3 Byzantine processes, and without using cryptography techniques. Moreover, this reduction algorithm uses a single instance of the underlying binary consensus, and tolerates message re-ordering by Byzantine processes. This work, done in collaboration with Achour Mostefaoui from LS2N (Nantes), appeared in Acta Informatica [20].

## **6.1.9.** A distributed leader election algorithm in crash-recovery and omissive system **Participant:** Michel Raynal.

We introduced a new distributed leader election algorithm for crash-recovery and omission environments. Contrary to previous works, our algorithm tolerates the occurrence of crash-recoveries and message omissions to any process during some finite but unknown time, after which a majority of processes in the system remains up and does not omit messages. This work, done in collaboration with Christian Fernández-Campusano, Mikel Larrea, and Roberto Cortiñas from UPV/EHU, Spain, appeared in Information Processing Letters 2017 [16].

#### 6.1.10. Providing Collision-Free and Conflict-Free Communication in General Synchronous Broadcast/Receive Networks

Participants: Michel Raynal, François Taïani.

This work [26] considers the problem of communication in dense and large scale wireless networks composed of resource-limited nodes. In this kind of networks, a massive amount of data is becoming increasingly available, and consequently implementing protocols achieving error-free communication channels constitutes

an important challenge. Indeed, in this kind of networks, the prevention of message conflicts and message collisions is a crucial issue. In terms of graph theory, solving this issue amounts to solve the distance-2 coloring problem in an arbitrary graph. The work presents a distributed algorithm providing the processes with such a coloring. This algorithm is itself collision-free and conflict-free. It is particularly suited to wireless networks composed of nodes with communication or local memory constraints.

This work was performed in collaboration with Abdelmadjid Bouabdallah and Hicham Lakhlef from Université Technologique de Compiègne (France).

## 6.1.11. Randomized abortable mutual exclusion with constant amortized RMR complexity on the CC model.

Participant: George Giakkoupis.

In [30], we presented an abortable mutual exclusion algorithm for the cache-coherent (CC) model with atomic registers and CAS objects. The algorithm has constant expected amortized RMR complexity in the oblivious adversary model and is deterministically deadlock-free. This is the first abortable mutual exclusion algorithm that achieves  $o(\log n / \log \log n)$  RMR complexity.

This work was done in collaboration with Philipp Woelfel (University of Calgary).

#### 6.2. Network and Graph Algorithms

#### 6.2.1. Tight bounds on vertex connectivity under sampling

Participant: George Giakkoupis.

A fundamental result by Karger (SODA 1994) states that for any  $\lambda$ -edge-connected graph with n nodes, independently sampling each edge with probability  $p = \Omega(\log(n)/\lambda)$  results in a graph that has edge connectivity  $\Omega(\lambda p)$ , with high probability. In [15], we proved the analogous result for vertex connectivity, when either vertices or edges are sampled. We showed that for any k-vertex-connected graph G with n nodes, if each node is independently sampled with probability  $p = \Omega(\sqrt{\log(n)/k})$ , then the subgraph induced by the sampled nodes has vertex connectivity  $\Omega(kp^2)$ , with high probability. If edges are sampled with probability  $p = \Omega(\log(n)/k)$ , then the subgraph induced by the sampled nodes has vertex connectivity  $\Omega(kp^2)$ , with high probability. If edges are sampled with probability  $p = \Omega(\log(n)/k)$ , then the sampled subgraph has vertex connectivity  $\Omega(kp)$ , with high probability. Both bounds are existentially optimal.

This work was done in collaboration with Keren Censor-Hillel (Technion), Mohsen Ghaffari (MIT), Bernhard Haeupler (Carnegie Mellon University), and Fabian Kuhn (University of Freiburg).

#### 6.2.2. Tight bounds for coalescing-branching random walks on regular graphs

#### Participant: George Giakkoupis.

A coalescing-branching random walk (Cobra) is a natural extension to the standard random walk on a graph. The process starts with one pebble at an arbitrary node. In each round of the process every pebble splits into k pebbles, which are sent to k random neighbors. At the end of the round all pebbles at the same node coalesce into a single pebble. The process is also similar to randomized rumor spreading, with each informed node pushing the rumor to k random neighbors each time it receives a copy of the rumor. Besides its mathematical interest, this process is relevant as an information dissemination primitive and a basic model for the spread of epidemics.

In [25] we studied the *cover time* of Cobra walks, which is the time until each node has seen at least one pebble. Our main result is a bound of  $O(\phi^{-1} \log n)$  rounds with high probability on the cover time of a Cobra walk with k = 2 on any regular graph with n nodes and conductance  $\phi$ . This bound improves upon all previous bounds in terms of graph expansion parameters. Moreover, we showed that for any connected regular graph the cover time is  $O(n \log n)$  with high probability, independently of the expansion. Both bounds are asymptotically tight.

This work was done in collaboration with Petra Berenbrink (University of Hamburg), Peter Kling (University of Hamburg).

#### **6.3. Scalable Systems**

#### 6.3.1. Agar: A Caching System for Erasure-Coded Data

Participants: Anne-Marie Kermarrec, François Taïani.

Erasure coding is an established data protection mechanism. It provides high resiliency with low storage overhead, which makes it very attractive to storage systems developers. Unfortunately, when used in a distributed setting, erasure coding hampers a storage system's performance, because it requires clients to contact several, possibly remote sites to retrieve their data. This has hindered the adoption of erasure coding in practice, limiting its use to cold, archival data. Recent research showed that it is feasible to use erasure coding for hot data as well, thus opening new perspectives for improving erasure-coded storage systems. In this work [32], we address the problem of minimizing access latency in erasure-coded storage. We propose Agar—a novel caching system tailored for erasure-coded content. Agar optimizes the contents of the cache based on live information regarding data popularity and access latency to different data storage sites. Our system adapts a dynamic programming algorithm to optimize the choice of data blocks that are cached, using an approach akin to "Knapsack " algorithms. We compare Agar to the classical Least Recently Used and Least Frequently Used cache eviction policies, while varying the amount of data cached between a data chunk and a whole replica of the object. We show that Agar can achieve 16% to 41% lower latency than systems that use classical caching policies.

This work was performed in collaboration with from Raluca Halalai and Pascal Felber from Université de Neuchâtel (Switzerland).

#### 6.3.2. Filament: A Cohort Construction Service for Decentralized Collaborative Editing Platforms

Participants: Resmi Ariyattu Chandrasekharannair, François Taïani.

Distributed collaborative editors allow several remote users to contribute concurrently to the same document. Only a limited number of concurrent users can be supported by the currently deployed editors. A number of peer-to-peer solutions have therefore been proposed to remove this limitation and allow a large number of users to work collaboratively. These approaches however tend to assume that all users edit the same set of documents, which is unlikely to be the case if such systems should become widely used and ubiquitous. In this work [24] we discuss a novel cohort-construction approach that allow users editing the same documents to rapidly find each other. Our proposal utilises the semantic relations between peers to construct a set of self-organizing overlays to route search requests. The resulting protocol is efficient, scalable, and provides beneficial load-balancing properties over the involved peers. We evaluate our approach and compare it against a standard Chord based DHT approach. Our approach performs as well as a DHT based approach but provides better load balancing.

#### 6.3.3. Scalable Anti-KNN: Decentralized Computation of k-Furthest-Neighbor Graphs with HyFN

Participants: Simon Bouget, David Bromberg, François Taïani.

The decentralized construction of k-Furthest-Neighbor graphs has been little studied, although such structures can play a very useful role, for instance in a number of distributed resource allocation problems. In this work [27] we define KFN graphs; we propose HyFN, a generic peer-to-peer KFN construction algorithm, and thoroughly evaluate its behavior on a number of logical networks of varying sizes. 1 Motivation k-Nearest-Neighbor (KNN) graphs have found usage in a number of domains, including machine learning, recommenders, and search. Some applications do not however require the k closest nodes, but the k most dissimilar nodes, what we term the k-Furthest-Neighbor (KFN) graph. Virtual Machines (VMs) placement —i.e. the (re-)assignment of workloads in virtualised IT environments— is a good example of where KFN can be applied. The problem consists in finding an assignment of VMs on physical machines (PMs) that minimises some cost function(s). The problem has been described as one of the most complex and important for the IT industry, with large potential savings. An important challenge is that a solution does not only consist

in packing VMs onto PMs — it also requires to limit the amount of interferences between VMs hosted on the same PM. Whatever technique is used (e.g. clustering), interference aware VM placement algorithms need to identify complementary workloads — i.e. workloads that are dissimilar enough that the interferences between them are minimised. This is why the application of KFN graphs would make a lot of sense: identifying quickly complementary workloads (using KFN) to help placement algorithms would decrease the risks of interferences. The construction of KNN graphs in decentralized systems has been widely studied in the past. However, existing approaches typically assume a form of "likely transitivity" of similarity between nodes: if A is close to B, and B to C, then A is likely to be close to C. Unfortunately this property no longer holds when constructing KFN graphs. As a result, these approaches are not working anymore when applied to this new problem.

This work was performed in collaboration with Anthony Ventresque from University College Dublin (Ireland).

#### 6.3.4. Density and Mobility-driven Evaluation of Broadcast Algorithms for MANETs

Participants: Simon Bouget, David Bromberg, François Taïani.

Broadcast is a fundamental operation in Mobile Ad-Hoc Networks (MANETs). A large variety of broadcast algorithms have been proposed. They differ in the way message forwarding between nodes is controlled, and in the level of information about the topology that this control requires. Deployment scenarios for MANETs vary widely, in particular in terms of nodes density and mobility. The choice of an algorithm depends on its expected coverage and energy cost, which are both impacted by the deployment context. In this work, we are interested in the comprehensive comparison of the costs and effectiveness of broadcast algorithms, representative of the main design alternatives. Our study reveals that the best algorithm for a given situation, such as a high density and a stable network, is not necessarily the most appropriate for a different situation such as a sparse and mobile network. We identify the algorithms characteristics that are correlated with these differences and discuss the pros and cons of each design.

This work was done in collaboration with Etienne Rivière (University of Neuchatel), Laurent Réveillère (University of Bordeaux) and appeared in ICDCS 2017

#### 6.3.5. An Adaptive Peer-Sampling Protocol for Building Networks of Browsers Participant: Davide Frey.

Peer-sampling protocols constitute a fundamental mechanism for a number of large-scale distributed applications. The recent introduction of WebRTC facilitated the deployment of decentralized applications over a network of browsers. However, deploying existing peer-sampling protocols on top of WebRTC raises issues about their lack of adaptiveness to sudden bursts of popularity over a network that does not manage addressing or routing. In this contribution, we introduced SPRAY, a novel random peer-sampling protocol that dynamically, quickly, and efficiently self-adapts to the network size. We evaluated SPRAY by means of simulations and real-world experiments. This demonstrated its flexibility and highlighted its efficiency improvements at the cost of small overhead. We embedded SPRAY in a real-time decentralized editor running in browsers and ran experiments involving up to 600 communicating web browsers. The results demonstrate that SPRAY significantly reduces the network traffic according to the number of participants and saves bandwidth.

This work was carried out in collaboration with Brice Nédelec, Julian Tanke, Pascal Molli, and Achour Mostéfaoui from the University of Nantes and will appear in the World Wide Web Journal [21].

#### 6.3.6. Designing Overlay Networks for Decentralized Clouds Participant: Marin Bertier.

Recent increase in demand for next-to-source data processing and low-latency applications has shifted attention from the traditional centralized cloud to more distributed models such as edge computing. In order to fully leverage these models it is necessary to decentralize not only the computing resources but also their management. While a decentralized cloud has various inherent advantages, it also introduces different challenges with respect to coordination and collaboration between resources. A large-scale system with

multiple administrative entities requires an overlay network which enables data and service localization based only on a partial view of the network. Numerous existing overlay networks target different properties but they are built in a generic context, without taking into account the specific requirements of a decentralized cloud. In this work [34], done in collaboration with G. Tato et C. Tedeschi from the Myriads project team, we identified some of these requirements and introduced Koala, a novel overlay network designed specifically to meet them.

## 7. Partnerships and Cooperations

#### 7.1. National Initiatives

#### 7.1.1. ANR project SocioPlug

Participants: Davide Frey, Anne-Marie Kermarrec, Pierre-Louis Roman, Francois Taiani.

SocioPlug is a collaborative ANR project involving Inria (ASAP team), the Univ. Nantes, and LIRIS (INSA Lyon and Univ. Claude Bernard Lyon). The project emerges from the observation that the features offered by the Web 2.0 or by social media do not come for free. Rather they bring the implicit cost of privacy. Users are more of less consciously selling personal data for services. SocioPlug aims to provide an alternative for this model by proposing a novel architecture for large-scale, user centric applications. Instead of concentrating information of cloud platforms owned by a few economic players, we envision services made possible by cheap low-end plug computers available in every home or workplace. This will make it possible to provide a high amount of transparency to users, who will be able to decide their own optimal balance between data sharing and privacy.

#### 7.1.2. DeSceNt CominLabs

Participants: Resmi Ariyattu Chandrasekharannair, Davide Frey, Michel Raynal, Francois Taiani.

The DeSceNt project aims to ease the writing of distributed programs on a federation of plug computers. Plug computers are a new generation of low-cost computers, such as Raspberry pi (25\$), VIA- APC (49\$), and ZERO Devices Z802 (75\$), which offer a cheap and readily available infrastructure to deploy domestic on-line software. Plug computers open the opportunity for everyone to create cheap nano-clusters of domestic servers, host data and services and federate these resources with their friends, colleagues, and families based on social links. More particularly we will seek in this project to develop novel decentralized protocols than can encapsulate the notion of privacy-preserving federation in plug-based infrastructures. The vision is to use these protocols to provide a programming toolkit that can support the convergent data types being developed by our partner GDD (Gestion de Données Distribuées) at Univ. Nantes.

#### 7.1.3. ANR Blanc project Displexity

Participants: George Giakkoupis, Anne-Marie Kermarrec, Michel Raynal.

The Displexity project started in 2011. The aim of this ANR project that also involves researchers from Paris and Bordeaux is to establish the scientific foundations for building up a consistent theory of computability and complexity for distributed computing. One difficulty to be faced by DISPLEXITY is to reconcile two non necessarily disjoint sub-communities, one focusing on the impact of temporal issues, while the other focusing on the impact of spatial issues on distributed algorithms.

#### 7.1.4. ANR project PAMELA

Participants: Davide Frey, George Giakkoupis, Francois Taiani.

PAMELA is a collaborative ANR project involving ASAP, Inria Lille, UMPC, Mediego and Snips. The project aims at developing machine learning theories and algorithms in order to learn local and personalized models from data distributed over networked infrastructures. This project seeks to provide first answers to modern information systems built by interconnecting many personal devices holding private user data in the search of personalized suggestions and recommendations. More precisely, we will focus on learning in a collaborative way with the help of neighbors in a network. We aim to lay the first blocks of a scientific foundation for these new types of systems, in effect moving from graphs of data to graphs of data and learned models. We argue that this shift is necessary in order to address the new constraints arising from the decentralization of information that is inherent to the emergence of big data. We will in particular focus on the question of learning under communication and privacy constraints. A significant asset of the project is the quality of its industrial partners, SNIPS and MEDIEGO, who bring in their expertise in privacy protection and distributed computing as well as use cases and datasets. They will contribute to translate this fundamental research effort into concrete outcomes by developing personalized and privacy-aware assistants able to provide contextualized recommendations on small devices and smartphones.

#### 7.1.5. ANR project OBrowser

Participants: David Bromberg, Davide Frey, Francois Taiani.

OBrowser is a collaborative ANR project involving Inria (ASAP team), the Univ. Nantes, the Bretagne Sud. University, and Orange. The project emerges from the vision of designing and deploying distributed application on millions of machines using web-enabled technologies without relying on a cloud or a central authority. OBrowser proposes to build collaborative applications through a decentralized execution environment composed of users' browsers that autonomously manages issues such as communication, naming, heterogeneity, and scalability. The introduction of browser-to-browser communication with WebRTC's Datachannel has made these scenarios closer, but today only experts can afford to tackle the technical challenges associated with large-scale browser-based deployments such as decentralized instant-messaging (Firechat) and Infrastructure-less Mission Critical Push To Talk. O'Browser aims to solve these challenges by means of a novel programming framework.

#### 7.1.6. ANR project DESCARTES

Participants: George Giakkoupis, Michel Raynal, Francois Taiani.

DESCARTES is a collaborative ANR project involving ASAP, Labri (U. Bordeaux), Lafia (U. Paris Diderot), Vérimag (Grenoble), LIF (Marseilles), and LINA (Nantes). Despite the practical interests of reusable frameworks for implementing specific distributed services, many of these frameworks still lack solid theoretical bases, and only provide partial solutions for a narrow range of services. In this project, we argue that this is mainly due to the lack of a generic framework that is able to unify the large body of fundamental knowledge on distributed computation that has been acquired over the last 40 years. The DESCARTES project aims at bridging this gap, by developing a systematic model of distributed computation that organizes the functionalities of a distributed computing system into reusable modular constructs assembled via well-defined mechanisms that maintain sound theoretical guarantees on the resulting system. DESCARTES arises from the strong belief that distributed computing is now mature enough to resolve the tension between the social needs for distributed computing systems, and the lack of a fundamentally sound and systematic way to realize these systems.

#### 7.1.7. ANR-ERC Tremplin project NDFUSION

Participant: George Giakkoupis.

NDFUSION is an 18-month ANR project awarded to the PI to support his preparation for his upcoming ERC grant application. The idea of intervening in a network diffusion process to enhance or retard its spread has been studied in various contexts, e.g., to increase the spread or speed of diffusion by choosing an appropriate set of seed nodes (a standard goal in viral marketing by word-of-mouth), or achieve the opposite effect

either by choosing a small set of nodes to remove (a goal in immunization against diseases), or by seeding a competing diffusion (e.g., to limit the spread of misinformation in a social network). The aim of this project is to consolidate existing work under a single, comprehensive framework, and using this framework to develop new, efficient algorithms for optimizing (maximizing or minimizing) the spread of diffusion processes. Novel aspects of the project involve issues of scalability, multiple concurrent diffusions, and the use of multistage online strategies to optimize diffusions. Results from this project are likely to be relevant to many different disciplines, from network optimization in computing to disease containment in medicine.

#### 7.2. International Initiatives

#### 7.2.1. Inria International Labs

• Anne-Marie Kermarrrec is the scientific co-chair (with Willy Zwaenepoel) of the EPFL/Inria International Lab

#### 7.2.2. Inria Associate Teams Not Involved in an Inria International Labs

7.2.2.1. LiDiCo

Title: Aux limites du calcul réparti

International Partner (Institution - Laboratory - Researcher):

UNAM (Mexico) - Instituto de Matematicas - Sergio Rajsbaum

Start year: 2017

See also: https://sites.google.com/site/lidicoequipeassociee/

Today distributed applications are pervasive, some very successful (e.g., Internet, P2P, social networks, cloud computing), and benefit everyone, but the design and the implementation of many of them still rely on ad-hoc techniques instead of on a solid theory. The next generation of distributed applications and services will be more and more complex and demands research efforts in establishing sound theoretical foundations to be able to master their design, their properties and their implementation. This proposal is a step in this inescapable direction.

#### 7.3. International Research Visitors

#### 7.3.1. Visits of International Scientists

- Peter Kling (U of Hamburg) visited ASAP (hosted by G Giakkoupis), Jan 19–25.
- Emanuele Natale (Max Planck, Saarbrücken) visited ASAP (hosted by G Giakkoupis), Apr 23–29.
- Thomas Sauerwald (U of Cambridge) visited ASAP (hosted by G Giakkoupis), Aug 21-24.
- Robert Elsässer (Salzburg U) visited ASAP (hosted by G Giakkoupis), Sep 25–29.

#### 7.3.1.1. Internships

- Jodi Spacek from University of British Columbia, Research internship from May 2017 until Aug 2017, supervised by David Bromberg.
- Stewart Grant from University of British Columbia, Research internship from May 2017 until Aug 2017, supervised by David Bromberg.
- Hayk Saribekyan from MIT, research Internship from June 2017 to August 2017.

#### 7.3.2. Visits to International Teams

7.3.2.1. Research Stays Abroad

- Michel Raynal was at the Hong Kong Polytechnic University from 15 September to 14 October 2017,
- David Bromberg did a visit at USP Department of Computer Science University of São Paulo, Sao Paulo, Brazil from February 22, 2017 to March 24, 2017

## 8. Dissemination

#### 8.1. Promoting Scientific Activities

#### 8.1.1. Scientific Events Organisation

#### 8.1.1.1. Member of the Organizing Committees

• George Giakkoupis was a co-organizer of the Workshop on Decentralized Machine Learning, Optimization and Privacy, Lille, France, Sep 2017.

#### 8.1.2. Scientific Events Selection

#### 8.1.2.1. Member of the Conference Program Committees

- François Taïani served on the TPC of the 18th ACM/IFIP/Usenix International Conference on Middleware (Middleware 2017).
- François Taïani served on the TPC of the 37th IEEE International Conference on Distributed Computing Systems (ICDCS 2017), in the Distributed Operating Systems and Middleware track.
- François Taïani served on the TPC of the 16th Workshop on Adaptive and Reflective Middleware (ARM 2017).
- François Taïani served on the TPC of the 1st Workshop on Scalable and Resilient Infrastructures for Distributed Ledgers (SERIAL 2017)
- George Giakkoupis served as a program committee member for the 31st IEEE International Parallel and Distributed Processing Symposium (IPDPS), Orlando, Florida, USA, May 2017.
- George Giakkoupis served as a program committee member for the 44th International Colloquium on Automata, Languages, and Programming (ICALP), Warsaw, Poland, July 2017.
- George Giakkoupis served as a program committee member for the 31st International Symposium on Distributed Computing (DISC), Vienna, Austria, Oct 2017.
- Davide Frey served as a program committee member for the 11th ACM International Conference on Distributed and Event-Based Systems (DEBS 2017), Barcelona, Spain, Jun 2017.
- Davide Frey served as a program committee member for the International Symposium on Computer Architecture and High Performance Computing, SBAC PAD 2017.
- Pierre-Louis Roman served in the program committee of the 7th IEEE International Symposium on Cloud and Service Computing (SC2 2017), Kanazawa, Japan, November 2017.
- Pierre-Louis Roman served in the program committee of the 10th IEEE International Conference on Service-Oriented Computing and Applications (SOCA 2017), Kanazawa, Japan, November 2017.

#### 8.1.3. Journal

#### 8.1.3.1. Member of the Editorial Boards

Anne-Marie Kermarrec is an associate editor of IEEE Internet Computing.

Anne-Marie Kermarrec is an associate editor of the Springer Computing Journal.

- 8.1.3.2. Reviewer Reviewing Activities
  - David Bromberg was reviewer for Journal of Internet Services and Applications
  - Davide Frey was a reviewer for PPNA Peer-to-Peer Networking and Applications.
  - Davide Frey was a reviewer for Transactions on Cloud Computing.
  - Davide Frey was a reviewer for Transactions on Sensor Networks.

#### 8.1.4. Invited Talks

• Anne-Marie Kermarrec gave a keynote talk at DISC 2017, Vienna, Austria

- Davide Frey gave an Invited talk at the "Session d'information sur les appels a` projets 2017 Big Data" at Business France, Paris, on January 6, 2017.
- David Bromberg gave an invited talk on "Dolmen: Towards the programmatic assembly of largescale distributed systems" as a UFG Seminar, UFG - Instituto de Informática, Goiania, Brazil, February 20, 2017
- David Bromberg gave an invited talk on "Dolmen: Towards the programmatic assembly of largescale distributed systems" as a UFG Seminar, Department of Computer Science, University of São Paulo, Sao Paulo, Brazil, February 22, 2017
- David Bromberg gave an invited talk on "Interoperability in distributed systems: past, present and future" at DISCOTEC/DIEBS, University of Neuchatel, June 21, University of Neuchatel, 2017

#### 8.1.5. Leadership within the Scientific Community

- François Taïani has been a member of the Steering Committee of IFIP WG 6.1 International Conference on Distributed Applications and Interoperable Systems (DAIS) since 2013.
- François Taïani has been a member of the Steering Committee of the ACM/IFIP/USENIX International Conference on Middleware (Middleware) since 2014.
- Anne-Marie Kermarrecc is a member of the ARCEP Prospective Board since 2015

#### 8.1.6. Research Administration

- François Taïani has been a member of the Scientific Orientation Committee (COS) of IRISA (UMR 6074) since 2013.
- Anne-Marie Kermarrrec was chair of the ERC Consolidator Grant Pannel in 2017
- Davide Frey is "correspondant scientifique Europe" at the DPEI for Inria Rennes.
- Davide Frey is an associate member of the COST-GTRI of Inria.
- David Bromberg is responsible of International relationships of IRISA
- David Bromberg is member of the scientific committee of the Media & Networks competitiveness cluster
- David Bromberg is scientific correspondent for ICT Digital in Rennes

#### 8.2. Teaching - Supervision - Juries

#### 8.2.1. Teaching

- Engineering School: David Bromberg, tools and programming languages for the Web, 48h, 2nd year of Engineering School (M1), ESIR / Université of Rennes 1, France.
- Engineering School: David Bromberg, Distributed software architecture, 24h, 2nd year of Engineering School (M1), ESIR / Université of Rennes 1, France.
- Pôle Universitaire Français: David Bromberg, Distributed software architecture, 60h, (M2), PUF / Université of Bordeaux, Vietnam.
- Engineering School: David Bromberg, Network security, 48h, 2nd year of Engineering School (M1), ESIR / Université of Rennes 1, France.
- Engineering School: David Bromberg, Cloud for the Internet of Things, 60h, 3rd year of Engineering School (M2), ESIR / Université of Rennes 1, France.
- Engineering School: David Bromberg, Internet of Things projects, 50h, 2rd year of Engineering School (M1), ESIR / Université of Rennes 1, France.
- University of Rennes 1: David Bromberg, Software engineering for the cloud, 4h, (M1), Université of Rennes 1, France.

- Engineering School: Francois Taiani, Synchronization and Parallel Programming, 48h, 2nd year of Engineering School (M1), ESIR / Univ. Rennes I, France.
- Engineering School: Francois Taiani, Distributed Systems, 24h, 3rd year of Engineering School (M2), ESIR / Univ. Rennes I, France.
- Engineering School: Francois Taiani, Parallel Algorithms for Big Data, 24h, 3rd year of Engineering School (M2), ESIR / Univ. Rennes I, France.
- Engineering School: Francois Taiani, Introduction to Operating Systems, 24h, 1st year of Engineering School (L3), ESIR / Univ. Rennes I, France.
- Master: Francois Taiani, Programming Technologies for the Cloud, 28h, M2, Univ. Rennes I, France.
- Master: Davide Frey, Scalable Distributed Systems, 10 hours, M1, EIT/ICT Labs Master School, Univ. Rennes I, France.
- Master: Davide Frey, Big-Data Storage and Processing Infrastructures, 10 hours, M2-SIF, Univ. Rennes I, France.

#### 8.2.2. Supervision

PhD: Stéphane Delbruel, Towards a Decentralized Embryomorphic Storage System [12], University of Rennes 1, 27 January 2017, Francois Taiani and Davide Frey.

PhD: Resmi Ariyattu Chandrasekharannair, Towards Decentralized Federations for Plug-based Decentralized Social Networks [11], University of Rennes 1, 5 July 2017, Francois Taiani.

PhD in progress : Pierre-Louis Roman, Epidemic Distributed Convergence for Decentralized Social Networking, Oct 2014, Francois Taiani and Davide Frey.

PhD in progress : Olivier Ruas, Dynamic Learning and Recommendations in Very Large Distributed Computing Infrastructures, Oct 2015, Anne-Marie Kermarrec and Francois Taiani.

PhD in progress: Simon Bouget, EMILIO: Emergent Middleware for Extra-Large-Scale Self Adaptation, Sep 2014, Francois Taiani and David Bromberg (since Sep 2015).

PhD in progress: Adrien Luxey, Towards New Solutions to Build Large Scale Distributed Applications in the Cloud, Oct 2016, David Bromberg.

PhD in progress: Louison Gitzinger, "Combattre les menaces à la confidentialité des données des utilisateurs sur Android.", Oct 2017, David Bromberg.

#### 8.2.3. Juries

- François Taïani served as external examiner on the PhD defense committee of Gowri Sankar Ramachandran on 12 June 2017 at KU Leuven (Belgium)
- François Taïani as served examiner on the PhD defense committee of Nicolas Keriven on 12 October 2017 at University of Rennes 1 (France)
- François Taïani served as reviewer on the PhD defense committee of Joachim Queireix on 21 November 2017 at Université de Bordeaux (France)
- François Taïani served as chair on the PhD defense committee of Florian Grandhomme on 23 November 2017 at Université de Rennes 1 (France)
- François Taïani served as chair on the PhD defense committee of Orçun Yildiz on 8 December 2017 at École Normale Supérieure de Rennes (France)
- Anne-Marie Kermarrec served as an examiner in the HDR defense of Christian Grothoff on October 18, 2017.

#### 8.3. Popularization

• Davide Frey published an article on the Telecom ParisTech Magazine. Issue 185.

• Pierre-Louis Roman gave an Invited talk at the Regional council of notaries of Brittany, Rennes, France, December 15, 2017. Title "Demystifying the blockchain and its utility for notaries" (talk in French).

## 9. Bibliography

#### Major publications by the team in recent years

- [1] F. ANDRÉ, A.-M. KERMARREC, E. LE MERRER, N. LE SCOUARNEC, G. STRAUB, A. VAN KEM-PEN.Archiving Cold Data in Warehouses with Clustered Network Coding, in "EuroSys 2014", Amsterdam, Netherlands, ACM New York, NY, USA, April 2014 [DOI: 10.1145/2592798.2592816], https://hal.inria. fr/hal-00994660.
- [2] M. BERTIER, D. FREY, R. GUERRAOUI, A.-M. KERMARREC, V. LEROY. The Gossple Anonymous Social Network, in "ACM/IFIP/USENIX 11th International Middleware Conference", India Bangalore, November 2010.
- [3] S. BOUGET, A.-M. KERMARREC, H. KERVADEC, F. TAÏANI. Polystyrene: the Decentralized Data Shape That Never Dies, in "The 2014 IEEE 34th International Conference on Distributed Computing Systems (ICDCS 2014)", Madrid, Spain, June 2014, p. 288 - 297 [DOI: 10.1109/ICDCS.2014.37], https://hal.inria.fr/hal-01080608.
- [4] J. CAO, M. RAYNAL, X. YANG, W. WU. Design and Performance Evaluation of Efficient Consensus Protocols for Mobile Ad Hoc Networks, in "IEEE Transactions on Computers", 2007, vol. 56, n<sup>o</sup> 8, p. 1055–1070.
- [5] K. CENSOR-HILLEL, M. GHAFFARI, G. GIAKKOUPIS, B. HAEUPLER, F. KUHN. *Tight Bounds on Vertex Connectivity Under Vertex Sampling*, in "26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2015)", San Diego, CA, United States, January 2015, p. 2006-1018 [DOI: 10.1137/1.9781611973730.133], https://hal.inria.fr/hal-01250519.
- [6] G. GIAKKOUPIS, P. WOELFEL. Randomized Mutual Exclusion with Constant Amortized RMR Complexity on the DSM, in "55th IEEE Symposium on Foundations of Computer Science (FOCS)", Philadelphia, PA, United States, October 2014, https://hal.inria.fr/hal-01097144.
- [7] M. JELASITY, S. VOULGARIS, R. GUERRAOUI, A.-M. KERMARREC, M. VAN STEEN. *Gossip-Based Peer Sampling*, in "ACM Transactions on Computer Systems", August 2007, vol. 41, n<sup>o</sup> 5.
- [8] A.-M. KERMARREC, F. TAÏANI, J. M. TIRADO MARTIN. Scaling Out Link Prediction with SNAPLE, in "16th Annual ACM/IFIP/USENIX Middleware Conference", Vancouver, Canada, December 2015, 12 [DOI: 10.1145/2814576.2814810], https://hal.inria.fr/hal-01244663.
- [9] A. MOSTEFAOUI, H. MOUMEN, M. RAYNAL.Signature-Free Asynchronous Binary Byzantine Consensus with t < n/3, O(n<sup>2</sup>) Messages, and O(1) Expected Time, in "Journal of the ACM (JACM)", December 2015, vol. 62, p. 1000-1020 [DOI : 10.1145/2785953], https://hal.archives-ouvertes.fr/hal-01176110.
- [10] F. TAÏANI, S. LIN, G. S. BLAIR.GOSSIPKIT: A Unified Component Framework for Gossip, in "IEEE Transactions on Software Engineering", March 2014, vol. 40, n<sup>o</sup> 2, p. 123-136 [DOI: 10.1109/TSE.2013.50], https://hal.inria.fr/hal-01080198.

#### **Publications of the year**

#### **Doctoral Dissertations and Habilitation Theses**

- [11] R. ARIYATTU. Towards federated social infrastructures for plug-based decentralized social networks, Université Rennes 1, July 2017, https://tel.archives-ouvertes.fr/tel-01622349.
- [12] S. DELBRUEL. Towards an architecture for tag-based predictive placement in distributed storage systems, Université Rennes 1, January 2017, https://tel.archives-ouvertes.fr/tel-01523568.

#### **Articles in International Peer-Reviewed Journal**

- [13] Z. BOUZID, M. RAYNAL, P. SUTRA. Anonymous obstruction-free (n,k)-set agreement with n-k+1 atomic read/write registers, in "Distributed Computing", May 2017, p. 1 19 [DOI: 10.1007/s00446-017-0301-7], https://hal.archives-ouvertes.fr/hal-01680833.
- [14] A. CASTAÑEDA, C. DELPORTE-GALLET, .. HUGUES FAUCONNIER, S. RAJSBAUM, M. RAYNAL. Making Local Algorithms Wait-Free: the Case of Ring Coloring, in "Theory of Computing Systems", May 2017, p. 1-22 [DOI: 10.1007/s00224-017-9772-Y], https://hal.archives-ouvertes.fr/hal-01672723.
- [15] K. CENSOR-HILLEL, M. GHAFFARI, G. GIAKKOUPIS, B. HAEUPLER, F. KUHN. Tight Bounds on Vertex Connectivity Under Sampling, in "ACM Transactions on Algorithms", May 2017, vol. 13, n<sup>o</sup> 2, p. 19:1 -19:26 [DOI : 10.1145/3086465], https://hal.inria.fr/hal-01635743.
- [16] C. FERNÁNDEZ-CAMPUSANO, M. LARREA, R. CORTIÑAS, M. RAYNAL. A distributed leader election algorithm in crash-recovery and omissive systems, in "Information Processing Letters", February 2017, vol. 118, p. 100 - 104 [DOI: 10.1016/J.IPL.2016.10.007], https://hal.inria.fr/hal-01660694.
- [17] R. GUERRAOUI, A.-M. KERMARREC, T. LIN, R. PATRA. Heterogeneous Recommendations: What You Might Like To Read After Watching Interstellar, in "Proceedings of the VLDB Endowment (PVLDB)", June 2017, vol. 10, n<sup>0</sup> 10, p. 1070 - 1081 [DOI : 10.14778/3115404.3115412], https://hal.inria.fr/hal-01660746.
- [18] M. HERLIHY, S. RAJSBAUM, M. RAYNAL, J. STAINER. From wait-free to arbitrary concurrent solo executions in colorless distributed computing, in "Theoretical Computer Science", June 2017, vol. 683, p. 1 - 21 [DOI: 10.1016/J.TCS.2017.04.007], https://hal.inria.fr/hal-01660566.
- [19] A. MOSTEFAOUI, M. PETROLIA, M. RAYNAL, C. JARD. Atomic Read/Write Memory in Signature-Free Byzantine Asynchronous Message-Passing Systems, in "Theory of Computing Systems", May 2017, vol. 60, n<sup>o</sup> 4, p. 677 - 694 [DOI: 10.1007/S00224-016-9699-8], https://hal.inria.fr/hal-01661008.
- [20] A. MOSTEFAOUI, M. RAYNAL.Signature-free asynchronous Byzantine systems: from multivalued to binary consensus with t < n/3, O(n 2) messages, and constant time, in "Acta Informatica", August 2017, vol. 54, n<sup>o</sup> 5, p. 501 520 [DOI: 10.1007/s00236-016-0269-Y], https://hal.inria.fr/hal-01660687.
- [21] B. NÉDELEC, J. TANKE, P. MOLLI, A. MOSTEFAOUI, D. FREY. An Adaptive Peer-Sampling Protocol for Building Networks of Browsers, in "World Wide Web", 2017, vol. 25, 1678 [DOI: 10.1007/s11280-017-0478-5], https://hal.inria.fr/hal-01619906.

[22] M. RAYNAL. *Distributed Universal Constructions: a Guided Tour*, in "Bulletin of the EATCS", 2017, vol. 121, p. 1-34, https://hal.inria.fr/hal-01661004.

#### **International Conferences with Proceedings**

- [23] F. ANDRÉ, A.-M. KERMARREC, N. LE SCOUARNEC. Accelerated Nearest Neighbor Search with Quick ADC, in "ICMR 2017 - ACM International Conference on Multimedia Retrieval", Bucarest, Romania, Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval, June 2017, p. 159-166, https://arxiv. org/abs/1704.07355 [DOI: 10.1145/3078971.3078992], https://hal.inria.fr/hal-01660730.
- [24] R. ARIYATTU, F. TAÏANI.*Filament: A Cohort Construction Service for Decentralized Collaborative Editing Platforms*, in "DAIS 2017 7th IFIP WG 6.1 International Conference on Distributed Applications and Interoperable Systems", Neuchâtel, Switzerland, June 2017, p. 146-16 [DOI: 10.1007/978-3-319-59665-5\_11], https://hal.inria.fr/hal-01617214.
- [25] P. BERENBRINK, G. GIAKKOUPIS, P. KLING. *Tight Bounds for Coalescing-Branching Random Walks on Regular Graphs*, in "SODA 2018 Proceedings of the 29th ACM-SIAM Symposium on Discrete Algorithms", New Orleans, United States, January 2018, https://hal.inria.fr/hal-01635757.
- [26] A. BOUABDALLAH, H. LAKHLEF, M. RAYNAL, F. TAÏANI. Providing Collision-Free and Conflict-Free Communication in General Synchronous Broadcast/Receive Networks, in "AINA 2017 - 31st IEEE International Conference on Advanced Information Networking and Applications", Taipei, Taiwan, March 2017, p. 399 - 406 [DOI: 10.1109/AINA.2017.39], https://hal.inria.fr/hal-01620356.
- [27] S. BOUGET, D. BROMBERG, F. TAÏANI, A. VENTRESQUE.Scalable Anti-KNN: Decentralized Computation of k-Furthest-Neighbor Graphs with HyFN, in "DAIS 2017 - 17th IFIP WG 6.1 International Conference on Distributed Applications and Interoperable Systems", Neuchâtel, Switzerland, L. Y. CHEN, H. REISER (editors), Springer, June 2017, p. 101-114 [DOI: 10.1007/978-3-319-59665-5\_7], https://hal.inria.fr/hal-01617211.
- [28] A. BOUTET, A.-M. KERMARREC, N. MITTAL, F. TAÏANI.KIFF: un algorithme de construction de graphe KNN générique, rapide et évolutif, in "ALGOTEL 2017 - 19èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Quiberon, France, May 2017, https://hal.archives-ouvertes.fr/hal-01518587.
- [29] A. CASTAÑEDA, Y. MOSES, M. RAYNAL, M. ROY.*Early Decision and Stopping in Synchronous Consensus:* A Predicate-Based Guided Tour, in "International Conference on Networked Systems (NETYS)", Marrakech, Morocco, Proceedings of NETYS 2017: the 5th International Conference on Networked Systems, May 2017, vol. 96, p. 167 - 221 [DOI : 10.1007/978-3-319-59647-1\_16], https://hal.archives-ouvertes.fr/hal-01559723.
- [30] G. GIAKKOUPIS, P. WOELFEL.Randomized Abortable Mutual Exclusion with Constant Amortized RMR Complexity on the CC Model, in "PODC 2017 - Proceedings of the 36th ACM Symposium on Principles of Distributed Computing", Washington, DC, United States, July 2017 [DOI: 10.1145/3087801.3087837], https://hal.inria.fr/hal-01635734.
- [31] R. GUERRAOUI, A.-M. KERMARREC, M. TAZIKI. *The Utility and Privacy Effects of a Click*, in "SIGIR 2017
  Thr 40th International ACM SIGIR Conference on Research and Development in Information Retrieval", Tokyo, Japan, Proceedings of the 40th International ACM SIGIR Conference on Research and Development

in Information Retrieval, ACM, August 2017, p. 665-674 [DOI: 10.1145/3077136.3080783], https://hal. inria.fr/hal-01660733.

- [32] R. HALALAI, P. FELBER, A.-M. KERMARREC, F. TAÏANI.Agar: A Caching System for Erasure-Coded Data, in "ICDCS 2017 - 37th IEEE International Conference on Distributed Computing Systems", Atlanta, GA, United States, June 2017, p. 1-11 [DOI: 10.1109/ICDCS.2017.97], https://hal.inria.fr/hal-01617146.
- [33] D. LIMA, H. MIRANDA, F. TAÏANI.Simulation of Partial Replication in Distributed Transactional Memory, in "2017 Wireless Days Conference", Porto, Portugal, March 2017, p. 54 - 59 [DOI: 10.1109/WD.2017.7918115], https://hal.inria.fr/hal-01620360.
- [34] G. TATO, M. BERTIER, C. TEDESCHI. Designing Overlay Networks for Decentralized Clouds, in "Future-Cloud 2017 - International Workshop on the Future of Cloud Computing and Cloud Services", Hong-Kong, China, December 2017, https://hal.inria.fr/hal-01634409.

#### **Conferences without Proceedings**

[35] A. CASTAÑEDA, M. RAYNAL, S. RAJSBAUM.Long-Lived Tasks, in "NETYS 2017 - 5th International Conference on NETworked sYStems", Marrakech, Morocco, Springer, May 2017, vol. 10299, p. 439-454 [DOI: 10.1007/978-3-319-59647-1\_32], https://hal.archives-ouvertes.fr/hal-01660646.

#### **Research Reports**

- [36] D. FREY, R. GUERRAOUI, A.-M. KERMARREC, M. MONOD.Live Streaming with Gossip, Inria Rennes Bretagne Atlantique; RR-9039, March 2017, n<sup>o</sup> RR-9039, https://hal.inria.fr/hal-01479885.
- [37] D. IMBS, A. MOSTEFAOUI, M. PERRIN, M. RAYNAL. Another Look at the Implementation of Read/write Registers in Crash-prone Asynchronous Message-Passing Systems (Extended Version), IRISA, Inria Rennes; LS2N-University of Nantes; Technion - Israel Institute of Technology, February 2017, https://arxiv.org/abs/ 1702.08176, https://hal.archives-ouvertes.fr/hal-01476201.
- [38] D. IMBS, A. MOSTEFAOUI, M. PERRIN, M. RAYNAL.Set-Constrained Delivery Broadcast: Definition, Abstraction Power, and Computability Limits, LIF, Université Aix-Marseille ; LINA-University of Nantes ; IMDEA Software Institute ; Institut Universitaire de France ; IRISA, Université de Rennes, June 2017, https:// arxiv.org/abs/1706.05267 , https://hal.archives-ouvertes.fr/hal-01540010.

#### **Other Publications**

[39] D. IMBS, A. MOSTEFAOUI, M. PERRIN, M. RAYNAL. Which Broadcast Abstraction Captures k-Set Agreement?, December 2017, https://arxiv.org/abs/1705.04835 - working paper or preprint, https://hal.archivesouvertes.fr/hal-01660664.

#### **References in notes**

- [40] M. AGUILERA. A Pleasant Stroll Through the Land of Infinitely Many Creatures, in "ACM SIGACT News, Distributed Computing Column", 2004, vol. 35, n<sup>o</sup> 2.
- [41] D. ANGLUIN.Local and Global Properties in Networks of Processes, in "Proc. 12th ACM Symposium on Theory of Computing (STOC'80)", 1980.

- [42] H. ATTIYA, A. BAR-NOY, D. DOLEV. Sharing Memory Robustly in Message-passing Systems, in "J. ACM", January 1995, vol. 42, n<sup>o</sup> 1, p. 124–142, http://doi.acm.org/10.1145/200836.200869.
- [43] K. BIRMAN, M. HAYDEN, O. OZKASAP, Z. XIAO, M. BUDIU, Y. MINSKY.*Bimodal Multicast*, in "ACM Transactions on Computer Systems", May 1999, vol. 17, n<sup>o</sup> 2, p. 41-88.
- [44] A. DEMERS, D. GREENE, C. HAUSER, W. IRISH, J. LARSON, S. SHENKER, H. STURGIS, D. SWINE-HART, D. TERRY. *Epidemic algorithms for replicated database maintenance*, in "Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC'87)", August 1987.
- [45] P. EUGSTER, S. HANDURUKANDE, R. GUERRAOUI, A.-M. KERMARREC, P. KOUZNETSOV. Lightweight Probabilistic Broadcast, in "ACM Transaction on Computer Systems", November 2003, vol. 21, n<sup>o</sup> 4.
- [46] M. MERRITT, G. TAUBENFELD. Computing Using Infinitely Many Processes, in "Proc. 14th Int'l Symposium on Distributed Computing (DISC'00)", 2000.
- [47] S. RATNASAMY, P. FRANCIS, M. HANDLEY, R. KARP, S. SHENKER. *A Scalable Content-Addressable Network*, in "Conference of the Special Interest Group on Data Communication (SIGCOMM'01)", 2001.
- [48] A. ROWSTRON, P. DRUSCHEL. *Pastry: Scalable, distributed object location and routing for large-scale peerto-peer systems*, in "IFIP/ACM Intl. Conf. on Distributed Systems Platforms (Middleware)", 2001.
- [49] I. STOICA, R. MORRIS, D. KARGER, F. KAASHOEK, H. BALAKRISHNAN. *Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications*, in "SIGCOMM'01", 2001.
- [50] S. VOULGARIS, D. GAVIDIA, M. VAN STEEN. CYCLON: Inexpensive Membership Management for Unstructured P2P Overlays, in "Journal of Network and Systems Management", 2005, vol. 13, n<sup>o</sup> 2.

# **Project-Team ASCOLA**

## Aspect and composition languages

IN COLLABORATION WITH: Laboratoire des Sciences du numerique de Nantes

IN PARTNERSHIP WITH: CNRS Ecole des Mines de Nantes IMT Atlantique Bretagne-Pays de la Loire Université Nantes

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Distributed programming and Software engineering

### **Table of contents**

1.	Personnel	. 36
2.	Overall Objectives	. 37
3.	Research Program	. 37
	3.1. Overview	37
	3.2. Software Composition	37
	3.3. Programming languages for advanced modularization	38
	3.4. Distribution and Concurrency	39
	3.5. Security	40
	3.6. Green IT	40
	3.7. Capacity Planning for Large Scale Distributed System	41
4.	Application Domains	. 41
	4.1. Enterprise Information Systems and Services	41
	4.2. Capacity Planning in Cloud, Fog and Edge Computing	42
	4.3. Pervasive Systems	42
5.	Highlights of the Year	. 43
	5.1.1. Remarkable results: research and third-party funding	43
	5.1.2. The future: the Gallinette and Stack teams	43
	5.1.3. Awards	43
6.	New Software and Platforms	. 44
	6.1. btrCloud	44
	6.2. SimGrid	44
	6.3. VMPlaces	45
_	6.4. ENOS	45
7.	New Results	. 45
	7.1. Cloud programming and management	45
	7.1.1. Cloud infrastructures	45
	7.1.1.1. Geo-distributed Clouds	46
	7.1.1.2. Cloud and HPC convergence	46
	7.1.1.3. Virtualization simulation	47
	7.1.2. Deployment and reconfiguration in the Cloud	47
	7.1.2.1. Initial deployment and placement strategies	4/
	7.1.2.2. Capacity planning and scheduling	48
	7.1.2.3. Reconfiguration and self-management	48
	7.2.1 Description	49
	7.2.1. Renewable energy	49
	7.2.2. Energy-aware consolidation and reconfiguration	50
	7.2.1 Souvity and privacy	51
	7.2.1.1. Security and privacy	51
	7.3.1.2. Drivery Aware Data Starses	51
	7.3.1.2. Privacy-Aware Data Storage.	51
	7.3.1.5. Accountability for Cloud applications.	51
	7.3.2. Johnware ucverophient and programming languages	52
	7.3.2.2. Cloud and HPC programming	52
8	<b>Bilatoral Contracts and Crants with Industry</b>	52 52
о. Q	Diateral Confides and Granes with muusiry	. 33
7.	0.1 Regional Initiatives	. 33
	9.1. Regional initiatives 0.1.1.1 CoMe(ACloud	53
	0.1.1.2 SyMeTRIC	53
	Since Syncerice	55

	9.2. National Initiatives	54
	9.2.1. CominLabs laboratory of excellence	54
	9.2.1.1. EPOC	54
	9.2.1.2. PrivGen	54
	9.2.2. ANR	54
	9.2.2.1. GRECO (ANR)	54
	9.2.2.2. KerStream (ANR)	55
	9.2.3. FSN	55
	9.2.3.1. Hosanna (FSN)	55
	9.2.3.2. Hydda (FSN)	55
	9.2.4. CPER	55
	9.2.5. Inria Project Labs	56
	9.2.6. InriaHub	56
	9.2.7. Fond d'amorçage IMT Industrie du Futur 2017	56
	9.2.8. Connect Talent	57
	9.3. European Initiatives	57
	9.3.1.1. CoqHoTT	57
	9.3.1.2. BigStorage	57
	9.4. International Initiatives	58
	9.5. International Research Visitors	58
10.	Dissemination	59
	10.1. Promoting Scientific Activities	59
	10.1.1. Scientific Events Organisation	59
	10.1.2. Scientific Events Selection	59
	10.1.2.1. Chair of Conference Program Committees	59
	10.1.2.2. Member of the Conference Program Committees	59
	10.1.3. Journal	59
	10.1.3.1. Member of the Editorial Boards	59
	10.1.3.2. Reviewer - Reviewing Activities	59
	10.1.4. Invited Talks	60
	10.1.5. Scientific Expertise	60
	10.1.6. Kesearch Administration	60
	10.2. Teaching - Supervision - Juries	61
	10.2.1. Supervision	61
11	10.2.2. Juries	01
11.	ыюнодгарпу	01

\_\_\_\_

#### **Project-Team ASCOLA**

*Creation of the Project-Team: 2009 January 01, end of the Project-Team: 2017 December 31* **Keywords:** 

#### **Computer Science and Digital Science:**

- A1.1.6. Cloud
- A1.1.8. Security of architectures
- A1.1.13. Virtualization
- A1.3. Distributed Systems
- A1.6. Green Computing
- A2.1. Programming Languages
- A2.1.1. Semantics of programming languages
- A2.1.2. Object-oriented programming
- A2.1.3. Functional programming
- A2.1.4. Aspect-oriented programming
- A2.1.6. Concurrent programming
- A2.1.7. Distributed programming
- A2.1.10. Domain-specific languages
- A2.1.11. Proof languages
- A2.2.1. Static analysis
- A2.4.2. Model-checking
- A2.4.3. Proofs
- A2.5. Software engineering
- A2.6.2. Middleware
- A2.6.3. Virtual machines
- A3.1.3. Distributed data
- A3.1.5. Control access, privacy
- A3.1.8. Big data (production, storage, transfer)
- A4.5. Formal methods for security
- A4.6. Authentication
- A4.7. Access control
- A4.8. Privacy-enhancing technologies
- A7.1. Algorithms
- A7.2. Logic in Computer Science

#### **Other Research Topics and Application Domains:**

- B3.1. Sustainable development
- B4.5. Energy consumption
- B4.5.1. Green computing
- B5.1. Factory of the future
- B6.1. Software industry
- B6.1.1. Software engineering
- B6.1.2. Software evolution, maintenance
- B6.5. Information systems

## 1. Personnel

#### **Research Scientists**

Ismael Cuadrado Cordero [IMT Atlantique, Researcher, until Aug 2017] Shadi Ibrahim [Inria, Researcher, from Apr 2017] Guillaume Munch [Inria, Researcher, until Apr 2017] Nicolas Tabareau [Inria, Researcher, until Apr 2017, HDR]

#### **Faculty Members**

Mario Südholt [Team leader, IMT Atlantique, Professor, HDR] Frederico Alvares de Oliveira Junior [IMT Atlantique, Associate Professor, temporary position, until Aug 2017] Julien Cohen [Univ de Nantes, Associate Professor, until Apr 2017] Pierre Cointe [IMT Atlantique, Professor, HDR] Hélène Coullon [IMT Atlantique, Associate Professor, HDR] Remi Douence [IMT Atlantique, Associate Professor, until Apr 2017, HDR] Herve Grall [IMT Atlantique, Associate Professor, until Apr 2017, HDR] Herve Grall [IMT Atlantique, Associate Professor, until Apr 2017] Adrien Lebre [Inria, Advanced Research Position until Sep 2017, IMT Atlantique Associate Professor, HDR] Thomas Ledoux [IMT Atlantique, Associate Professor] Jean-Marc Menaud [IMT Atlantique, Professor, HDR] Jacques Noyé [IMT Atlantique, Associate Professor] Jean-Claude Royer [IMT Atlantique, Professor, HDR]

#### **Technical Staff**

Ronan-Alexandre Cherrueau [Inria] Remy Pottier [IMT Atlantique, until Aug 2017] Alexandre Van Kempen [Inria, from Oct 2017]

#### **PhD Students**

Mohamed Abderrahim [Orange Labs] Mohammad-Mahdi Bazm [Orange Labs] Paul Blouet [Ecole des Mines de Nantes, until Apr 2017] Fatima-Zahra Boujdad [Univ de Nantes] Simon Boulier [Ecole normale supérieure de Rennes, until Apr 2017] Emile Cadorel [Armines, from Oct 2017] Maverick Chardet [Inria, from Oct 2017] Bastien Confais [CNRS] David Espinel [Orange Labs, from Oct 2017] Gaetan Gilbert [IMT Atlantique, until Apr 2017] Md Sabbir Hasan [INSA, co-supervision with Myriads team, Inria, until Mar 2017] Yacine Hebbal [Orange Labs, until Oct 2017] Ambroise Lafont [IMT Atlantique, until Apr 2017] Thuy Linh Nguyen [Inria] Kevin Quirin [Ecole normale supérieure de Rennes, until Aug 2017] Igor Zhirkov [Armines]

#### **Post-Doctoral Fellows**

Benedikt Ahrens [Inria, until Apr 2017] Zakarea Al Shara [IMT Atlantique, until Oct 2017] Eric Finster [Inria, from Feb 2017 until Apr 2017] Jonathan Pastor [Ecole des Mines de Nantes , from Aug 2017] Dimitri Pertin [Inria] Anthony Simonet [Inria, until Jun 2017] Chi Zhou [Inria, from Apr 2017 until Aug 2017]
Pierre-Marie Pedrot [IMT Atlantique, until Apr 2017]

#### Visiting Scientist

Amin Timany [IMT Atlantique, from Mar 2017 until Apr 2017]

#### Administrative Assistant

Anne Claire Binetruy [Inria]

## 2. Overall Objectives

#### **2.1. Presentation**

The research team addresses the general problem of evolving software by developing concepts, languages, implementations and tools for building software architectures based on components and aspects. Its long term goal is the development of new abstractions for the programming of software architectures, their representation in terms of expressive programming languages and their correct and efficient implementation.

We pursue the following objectives:

- New concepts and techniques for the compositional definition and implementation of complex software systems, notably involving crosscutting concerns that cannot be handled modularly using traditional software development approaches.
- New programming techniques and algorithms for resource management in mutualized environments. We provide language abstractions and implementation techniques for large-scale applications in cloud- and grid-based systems, both on the level of (service-based) applications and (virtualized) infrastructures. We develop solutions, in particular, for the optimization of the energy consumption in such environments (data centers ...)
- We develop new formal theories for and apply formal methods to the correctness of software systems. We aim at developing more powerful techniques for theorem proving and enable complex, often dynamic, software systems to be proven correct using program transformations and analysis techniques. We develop solutions, in particular, for the constructive enforcement of security properties on the level of software systems.

Finally, we apply and validate our results based on real-world applications from numerous domains, notably enterprise information systems, the Cloud, and pervasive systems.

## **3. Research Program**

#### 3.1. Overview

Since we mainly work on new concepts for the language-based definition and implementation of complex software systems, we first briefly introduce some basic notions and problems of software components (understood in a broad sense, that is, including modules, objects, architecture description languages and services), aspects, and domain-specific languages. We conclude by presenting the main issues related to distribution and concurrency, in particular related to capacity planning issues that are relevant to our work.

#### **3.2. Software Composition**

**Modules and services.** The idea that building *software components*, i.e., composable prefabricated and parameterized software parts, was key to create an effective software industry was realized very early [77]. At that time, the scope of a component was limited to a single procedure. In the seventies, the growing complexity of software made it necessary to consider a new level of structuring and programming and led to the notions of information hiding, *modules*, and module interconnection languages [84], [60]. Information hiding promotes a black-box model of program development whereby a module implementation, basically a collection of procedures, is strongly encapsulated behind an interface. This makes it possible to guarantee logical invariant *properties* of the data managed by the procedures and, more generally, makes *modular reasoning* possible.

In the context of today's Internet-based information society, components and modules have given rise to *software services* whose compositions are governed by explicit *orchestration or choreography* specifications that support notions of global properties of a service-oriented architecture. These horizontal compositions have, however, to be frequently adapted dynamically. Dynamic adaptations, in particular in the context of software evolution processes, often conflict with a black-box composition model either because of the need for invasive modifications, for instance, in order to optimize resource utilization or modifications to the vertical compositions implementing the high-level services.

**Object-Oriented Programming.** Classes and objects provide another kind of software component, which makes it necessary to distinguish between *component types* (classes) and *component instances* (objects). Indeed, unlike modules, objects can be created dynamically. Although it is also possible to talk about classes in terms of interfaces and implementations, the encapsulation provided by classes is not as strong as the one provided by modules. This is because, through the use of inheritance, object-oriented languages put the emphasis on *incremental programming* to the detriment of modular programming. This introduces a white-box model of software development and more flexibility is traded for safety as demonstrated by the *fragile base class* issue [80].

Architecture Description Languages. The advent of distributed applications made it necessary to consider more sophisticated connections between the various building blocks of a system. The *software architecture* [89] of a software system describes the system as a composition of *components* and *connectors*, where the connectors capture the *interaction protocols* between the components [48]. It also describes the rationale behind such a given architecture, linking the properties required from the system to its implementation. *Architecture Description Languages* (ADLs) are languages that support architecture-based development [78]. A number of these languages make it possible to generate executable systems from architectural descriptions, provided implementations for the primitive components are available. However, guaranteeing that the implementation conforms to the architecture is an issue.

**Protocols.** Today, protocols constitute a frequently used means to precisely define, implement, and analyze contracts, notably concerning communication and security properties, between two or more hardware or software entities. They have been used to define interactions between communication layers, security properties of distributed communications, interactions between objects and components, and business processes.

Object interactions [82], component interactions [95], [86] and service orchestrations [61] are most frequently expressed in terms of *regular interaction protocols* that enable basic properties, such as compatibility, substitutability, and deadlocks between components to be defined in terms of basic operations and closure properties of finite-state automata. Furthermore, such properties may be analyzed automatically using, e.g., model checking techniques [58], [67].

However, the limited expressive power of regular languages has led to a number of approaches using more expressive *non-regular* interaction protocols that often provide distribution-specific abstractions, e.g., session types [71], or context-free or turing-complete expressiveness [87], [55]. While these protocol types allow conformance between components to be defined (e.g., using unbounded counters), property verification can only be performed manually or semi-automatically.

#### 3.3. Programming languages for advanced modularization

The main driving force for the structuring means, such as components and modules, is the quest for clean *separation of concerns* [62] on the architectural and programming levels. It has, however, early been noted that concern separation in the presence of crosscutting functionalities requires specific language and implementation level support. Techniques of so-called *computational reflection*, for instance, Smith's 3-Lisp or Kiczales's CLOS meta-object protocol [90], [74] as well as metaprogramming techniques have been developed to cope with this problem but proven unwieldy to use and not amenable to formalization and property analysis due to their generality. Methods and techniques from two fields have been particularly useful in addressing such advanced modularization problems: Aspect-Oriented Software Development as the field concerned with the systematic handling of modularization issues and domain-specific languages that provide declarative and efficient means for the definition of crosscutting functionalities.

**Aspect-Oriented Software Development** [73], [46] has emerged over the previous decade as the domain of systematic exploration of crosscutting concerns and corresponding support throughout the software development process. The corresponding research efforts have resulted, in particular, in the recognition of *crosscutting* as a fundamental problem of virtually any large-scale application, and the definition and implementation of a large number of aspect-oriented models and languages.

However, most current aspect-oriented models, notably AspectJ [72], rely on pointcuts and advice defined in terms of individual execution events. These models are subject to serious limitations concerning the modularization of crosscutting functionalities in distributed applications, the integration of aspects with other modularization mechanisms such as components, and the provision of correctness guarantees of the resulting AO applications. They do, in particular, only permit the manipulation of distributed applications on a perhost basis, that is, without direct expression of coordination properties relating different distributed entities [91]. Similarly, current approaches for the integration of aspects and (distributed) components do not directly express interaction properties between sets of components but rather seemingly unrelated modifications to individual components [59]. Finally, current formalizations of such aspect models are formulated in terms of low-level semantic abstractions (see, e.g., Wand's et al semantics for AspectJ [94]) and provide only limited support for the analysis of fundamental aspect properties.

Different approaches have been put forward to tackle these problems, in particular, in the context of socalled *stateful* or *history-based aspect languages* [63], [64], which provide pointcut and advice languages that directly express rich relationships between execution events. Such languages have been proposed to directly express coordination and synchronization issues of distributed and concurrent applications [83], [53], [66], provide more concise formal semantics for aspects and enable analysis of their properties [49], [65], [63], [47]. Furthermore, first approaches for the definition of *aspects over protocols* have been proposed, as well as over regular structures [63] and non-regular ones [93], [81], which are helpful for the modular definition and verification of protocols over crosscutting functionalities.

They represent, however, only first results and many important questions concerning these fundamental issues remain open, in particular, concerning the semantics foundations of AOP and the analysis and enforcement of correctness properties governing its, potentially highly invasive, modifications.

**Domain-specific languages (DSLs)** represent domain knowledge in terms of suitable basic language constructs and their compositions at the language level. By trading generality for abstraction, they enable complex relationships among domain concepts to be expressed concisely and their properties to be expressed and formally analyzed. DSLs have been applied to a large number of domains; they have been particularly popular in the domain of software generation and maintenance [79], [97].

Many modularization techniques and tasks can be naturally expressed by DSLs that are either specialized with respect to the type of modularization constructs, such as a specific brand of software component, or to the compositions that are admissible in the context of an application domain that is targeted by a modular implementation. Moreover, software development and evolution processes can frequently be expressed by transformations between applications implemented using different DSLs that represent an implementation at different abstraction levels or different parts of one application.

Functionalities that crosscut a component-based application, however, complicate such a DSL-based transformational software development process. Since such functionalities belong to another domain than that captured by the components, different DSLs should be composed. Such compositions (including their syntactic expression, semantics and property analysis) have only very partially been explored until now. Furthermore, restricted composition languages and many aspect languages that only match execution events of a specific domain (e.g., specific file accesses in the case of security functionality) and trigger only domain-specific actions clearly are quite similar to DSLs but remain to be explored.

#### 3.4. Distribution and Concurrency

While ASCOLA does not investigate distribution and concurrency as research domains per se (but rather from a software engineering and modularization viewpoint), there are several specific problems and corresponding

approaches in these domains that are directly related to its core interests that include the structuring and modularization of large-scale distributed infrastructures and applications. These problems include crosscutting functionalities of distributed and concurrent systems, support for the evolution of distributed software systems, and correctness guarantees for the resulting software systems.

Underlying our interest in these domains is the well-known observation that large-scale distributed applications are subject to *numerous crosscutting functionalities* (such as the transactional behavior in enterprise information systems, the implementation of security policies, and fault recovery strategies). These functionalities are typically partially encapsulated in distributed infrastructures and partially handled in an ad hoc manner by using infrastructure services at the application level. Support for a more principled approach to the development and evolution of distributed software systems in the presence of crosscutting functionalities has been investigated in the field of *open adaptable middleware* [54], [76]. Open middleware design exploits the concept of reflection to provide the desired level of configurability and openness. However, these approaches are subject to several fundamental problems. One important problem is their insufficient, framework-based support that only allows partial modularization of crosscutting functionalities.

There has been some *criticism* on the use of *AspectJ-like aspect models* (which middleware aspect models like that of JBoss AOP are an instance of) for the modularization of distribution and concurrency related concerns, in particular, for transaction concerns [75] and the modularization of the distribution concern itself [91]. Both criticisms are essentially grounded in AspectJ's inability to explicitly represent sophisticated relationships between execution events in a distributed system: such aspects therefore cannot capture the semantic relationships that are essential for the corresponding concerns. History-based aspects, as those proposed by the ASCOLA project-team provide a starting point that is not subject to this problem.

From a point of view of language design and implementation, aspect languages, as well as domain specific languages for distributed and concurrent environments share many characteristics with existing distributed languages: for instance, event monitoring is fundamental for pointcut matching, different synchronization strategies and strategies for code mobility [69] may be used in actions triggered by pointcuts. However, these relationships have only been explored to a small degree. Similarly, the formal semantics and formal properties of aspect languages have not been studied yet for the distributed case and only rudimentarily for the concurrent one [49], [66].

#### 3.5. Security

Security properties and policies over complex service-oriented and standalone applications become ever more important in the context of asynchronous and decentralized communicating systems. Furthermore, they constitute prime examples of crosscutting functionalities that can only be modularized in highly insufficient ways with existing programming language and service models. Security properties and related properties, such as accountability properties, are therefore very frequently awkward to express and difficult to analyze and enforce (provided they can be made explicit in the first place).

Two main issues in this space are particularly problematic from a compositional point of view. First, information flow properties of programming languages, such as flow properties of Javascript [51], and service-based systems [57] are typically specially-tailored to specific properties, as well as difficult to express and analyze. Second, the enforcement of security properties and security policies, especially accountability-related properties [85], [92], is only supported using ad hoc means with rudimentary support for property verification.

The ASCOLA team has recently started to work on providing formal methods, language support and implementation techniques for the modular definition and implementation of information flow properties as well as policy enforcement in service-oriented systems as well as, mostly object-oriented, programming languages.

#### 3.6. Green IT

With the emergence of the Future Internet and the dawn of new IT architecture and computation models such as cloud computing, the usage of data centers (DC) as well as their power consumption increase dramatically

[56]. Besides the ecological impact [70], energy consumption is a predominant criterion for DC providers since it determines the daily cost of their infrastructure. As a consequence, power management becomes one of the main challenges for DC infrastructures and more generally for large-scale distributed systems.

To address this problem, we study two approaches: a workload-driven [52] and power-driven one [88]. As part of the workload-driven solution, we adapt the power consumption of the DC depending on the application workload, and evaluate whether this workload to be more reactive. We develop a distributed system from the system to the service-oriented level mainly based on hardware and virtualization capabilities that is managed in a user-transparent fashion. As part of the power-driven approach, we address energy consumption issues through a strong synergy inside the infrastructure software stack and more precisely between applications and resource management systems. This approach is characterized by adapting QoS properties aiming at the best trade-off between cost of energy (typically from the regular electric grid), its availability (for instance, from renewable energy), and service degradation caused, for instance, by application reconfigurations to jobs suspensions.

#### 3.7. Capacity Planning for Large Scale Distributed System

Since the last decade, cloud computing has emerged as both a new economic model for software (provision) and as flexible tools for the management of computing capacity [50]. Nowadays, the major cloud features have become part of the mainstream (virtualization, storage and software image management) and the big market players offer effective cloud-based solutions for resource pooling. It is now possible to deploy virtual infrastructures that involve virtual machines (VMs), middleware, applications, and networks in such a simple manner that a new problem has emerged since 2010: VM sprawl (virtual machine proliferation) that consumes valuable computing, memory, storage and energy resources, thus menacing serious resource shortages. Scientific approaches that address VM sprawl are both based on classical administration techniques like the lifecycle management of a large number of VMs as well as the arbitration and the careful management of all resources consumed and provided by the hosting infrastructure (energy, power, computing, memory, network etc.) [68], [96].

The ASCOLA team investigates fundamental techniques for cloud computing and capacity planning, from infrastructures to the application level. Capacity planning is the process of planning for, analyzing, sizing, managing and optimizing capacity to satisfy demand in a timely manner and at a reasonable cost. Applied to distributed systems like clouds, a capacity planning solution must mainly provide the minimal set of resources necessary for the proper execution of the applications (i.e., to ensure SLA). The main challenges in this context are: scalability, fault tolerance and reactivity of the solution in a large-scale distributed system, the analysis and optimization of resources to minimize the cost (mainly costs related to the energy consumption of datacenters), as well as the profiling and adaptation of applications to ensure useful levels of quality of service (throughput, response time, availability etc.).

Our solutions are mainly based on virtualized infrastructures that we apply from the IaaS to the SaaS levels. We are mainly concerned by the management and the execution of the applications by harnessing virtualization capabilities, the investigation of alternative solutions that aim at optimizing the trade-off between performance and energy costs of both applications and cloud resources, as well as arbitration policies in the cloud in the presence of energy-constrained resources.

## 4. Application Domains

#### 4.1. Enterprise Information Systems and Services

Large IT infrastructures typically evolve by adding new third-party or internally-developed components, but also frequently by integrating already existing information systems. Integration frequently requires the addition of glue code that mediates between different software components and infrastructures but may also consist in more invasive modifications to implementations, in particular to implement crosscutting functionalities. In more abstract terms, enterprise information systems are subject to structuring problems involving horizontal composition (composition of top-level functionalities) as well as vertical composition (reuse and sharing of implementations among several top-level functionalities). Moreover, information systems have to be more and more dynamic.

Service-Oriented Computing (SOC) that is frequently used for solving some of the integration problems discussed above. Indeed, service-oriented computing has two main advantages:

- Loose-coupling: services are autonomous: they do not require other services to be executed;
- Ease of integration: Services communicate over standard protocols.

Our current work is based on the following observation: similar to other compositional structuring mechanisms, SOAs are subject to the problem of crosscutting functionalities, that is, functionalities that are scattered and tangled over large parts of the architecture and the underlying implementation. Security functionalities, such as access control and monitoring for intrusion detection, are a prime example of such a functionality in that it is not possible to modularize security issues in a well-separated module. Aspect-Oriented Software Development is precisely an application-structuring method that addresses in a systemic way the problem of the lack of modularization facilities for crosscutting functionalities.

We are considering solutions to secure SOAs by providing an aspect-oriented structuring and programming model that allows security functionalities to be modularized. Two levels of research have been identified:

- Service level: as services can be composed to build processes, aspect weaving will deal with the orchestration and the choreography of services.
- Implementation level: as services are abstractly specified, aspect weaving will require to extend service interfaces in order to describe the effects of the executed services on the sensitive resources they control.

#### 4.2. Capacity Planning in Cloud, Fog and Edge Computing

Cloud and more recently Fog and Edge computing platforms aim at delivering large capacities of computing power. These capacities can be used to improve performance (for scientific applications) or availability (e.g., for Internet services hosted by datacenters). These distributed infrastructures consist of a group of coupled computers that work together and may be spread across a LAN (cluster), across a the Internet (Fog/Edge). Due to their large scale, these architectures require permanent adaptation, from the application to the system level and call for automation of the corresponding adaptation processes. We focus on self-configuration and self-optimization functionalities across the whole software stack: from the lower levels (systems mechanisms such as distributed file systems for instance) to the higher ones (i.e. the applications themselves such as clustered servers or scientific applications).

In 2017, we have been consolidating our expertise around the OpenStack ecosystem. We proposed in particular EnOS, a dedicated framework to conduct performance analyses of OpenStack at large-scale ina reproducible manner. The framework enables researchers to conduct experiments in an automize manner on top of different testbeds such as Grid'5000 and Chameleon. see Sec. 7.1.

In the energy field, we have designed a set of techniques, named Optiplace, for cloud management with flexible power models through constraint programming. OptiPlace supports external models, named views. Specifically, we have developed a power view, based on generic server models, to define and reduce the power consumption of a datacenter's physical servers. We have shown that OptiPlace behaves at least as good as our previous system, Entropy, requiring as low as half the time to find a solution for the constrained-based placement of tasks for large datacenters.

#### 4.3. Pervasive Systems

Pervasive systems are another class of systems raising interesting challenges in terms of software structuring. Such systems are highly concurrent and distributed. Moreover, they assume a high-level of mobility and context-aware interactions between numerous and heterogeneous devices (laptops, PDAs, smartphones, cameras, electronic appliances...). Programming such systems requires proper support for handling various interfering concerns like software customization and evolution, security, privacy, context-awareness... Additionally, service composition occurs spontaneously at runtime.

Like Pervasive systems, Internet of thing is a major theme of these last ten years. Many research works has been led on the whole chain, from communicating sensors to big data management, through communication middlewares.

The more a sensor networks senses various data, the more the users panel is heterogeneous. Such an heterogeneity leads to a major problem about data modeling: for each user, to aim at precisely addressing his needs and his needs only; ie to avoid a data representation which would overwhelm the user with all the data sensed from the network, regardless if he needs it or not. To leverage this issue, we propose a multitree modeling for sensor networks which addresses each of these specific usages. With this modeling comes a domain specific language (DSL) which allows users to manipulate, parse and aggregate information from the sensors.

## 5. Highlights of the Year

#### 5.1. Highlights of the Year

#### 5.1.1. Remarkable results: research and third-party funding

Regarding scientific results, the team has produced a number of outstanding results on Fog/Edge architectures, notably on how to leverage renewable energy in this context [29], [9], [8], [33]. In the software engineering domain, particularly notable contributions have been made on software adaptability [4], [11].

Concerning third-party funding, 2017 has seen the acceptance of the large industrial/academic Hydda project as well as the start of two individual projects, the Kerdata and ConnectTalent projects, both of which issue of highly-competitive calls.

#### 5.1.2. The future: the Gallinette and Stack teams

After a 10-year adventure, the research path of the Ascola team finishes at the end of 2017 after having given rise to two new teams in 2017: the Gallinette team in April and the Stack team in November. These new teams pursue and diversify Ascola's main research domains, respectively formal methods for programming languages and distributed software systems. Note that because of the rather early split of the Gallinette team, we have not included the corresponding results in this year's Ascola report.

#### 5.1.3. Awards

In 2017 members of the team have been awarded three research-related awards: two personal awards and a best paper award:

#### • Programme Jeunes Talents France Chine 2017:

Shadi Ibrahim was one of the 12 researchers selected for the "Programme Jeunes Talents France Chine" award (12 out of 54 applicants).

#### • ICA3PP-2017 Outstanding Leadership Award:

Shadi Ibrahim received an Outstanding Leadership Award as program chair of the ICA3PP-2017. BEST PAPERS AWARDS :

## [27] The 7th International Conference on Cloud Computing and Services Science (CLOSER 2017). J. LEJEUNE, F. ALVARES, T. LEDOUX.

## 6. New Software and Platforms

#### 6.1. btrCloud

KEYWORDS: Cloud computing - Data center - Cluster - Placement - Autonomic system - Orchestration - Energy - Grid - Virtualization - Scheduler

FUNCTIONAL DESCRIPTION: Orchestration, virtualization, energy, autonomic system, placement, cloud computing, cluster, data center, scheduler, grid

btrCloud is a virtual machine manager for clusters and provides a complete solution for the management and optimization of virtualized data centers. btrCloud (acronym of better cloud) is composed of three parts.

The analysis function enables operatives and people in charge to monitor and analyze how a data-center works - be it on a daily basis, on the long run, or in order to predict future trends. This feature includes boards for performance evaluation and analysis as well as trends estimation.

btrCloud, by the integration of btrScript, provides (semi-)automated VM lifecycle management, including provisioning, resource pool management, VM tracking, cost accounting, and scheduled deprovisioning. Key features include a thin client interface, template-based provisioning, approval workflows, and policy-based VM placement.

Finally, several kinds of optimizations are currently available, such as energy and load balancing. The former can help save up to around 20

- Participants: Frédéric Dumont, Guillaume Le Louët and Jean-Marc Menaud
- Contact: Guillaume Le Louët
- URL: http://www.btrcloud.org/btrCloud/index\_EN.html

#### 6.2. SimGrid

KEYWORDS: Large-scale Emulators - Grid Computing - Distributed Applications

SCIENTIFIC DESCRIPTION: SimGrid is a toolkit that provides core functionalities for the simulation of distributed applications in heterogeneous distributed environments. The simulation engine uses algorithmic and implementation techniques toward the fast simulation of large systems on a single machine. The models are theoretically grounded and experimentally validated. The results are reproducible, enabling better scientific practices.

Its models of networks, cpus and disks are adapted to (Data)Grids, P2P, Clouds, Clusters and HPC, allowing multi-domain studies. It can be used either to simulate algorithms and prototypes of applications, or to emulate real MPI applications through the virtualization of their communication, or to formally assess algorithms and applications that can run in the framework.

The formal verification module explores all possible message interleavings in the application, searching for states violating the provided properties. We recently added the ability to assess liveness properties over arbitrary and legacy codes, thanks to a system-level introspection tool that provides a finely detailed view of the running application to the model checker. This can for example be leveraged to verify both safety or liveness properties, on arbitrary MPI code written in C/C++/Fortran.

**RELEASE FUNCTIONAL DESCRIPTION:** 

- Four releases in 2017. Major changes:
  - S4U: many progress, toward SimGrid v4.0. About 80% of the features offered by SimDag and MSG are now integrated, along with examples. Users can now write plugins to extend SimGrid.
  - SMPI: Support MPI 2.2, RMA support, Convert internals to C++.
  - Java: Massive memleaks and performance issues fixed.
  - New models: Multi-core VMs, Energy consumption due to the network

- All internals are now converted to C++, and most of our internally developped data containers were replaced with std::\* constructs.
- (+ bug fixes, cleanups and documentation improvements)
- Participants: Adrien Lèbre, Arnaud Legrand, Augustin Degomme, Florence Perronnin, Frédéric Suter, Jean-Marc Vincent, Jonathan Pastor, Jonathan Rouzaud-Cornabas, Luka Stanisic, Mario Südholt and Martin Quinson
- Partners: CNRS ENS Rennes
- Contact: Martin Quinson
- URL: http://simgrid.gforge.inria.fr/

#### **6.3. VMPlaces**

FUNCTIONAL DESCRIPTION: VMPlaces is a dedicated framework to evaluate and compare VM placement algorithms. This framework is composed of two major components: the injector and the VM placement algorithm. The injector is the generic part of the framework (i.e. the one you can directly use) while the VM placement algorithm is the part you want to study (or compare with available algorithms). Currently, the VMPlaceS is released with three algorithms:

Entropy, a centralized approach using a constraint programming approach to solve the placement/reconfiguration VM problem

Snooze, a hierarchical approach where each manager of a group invokes Entropy to solve the placement/reconfiguration VM problem. Note that in the original implementation of Snooze, it is using a specific heuristic to solve the placement/reconfiguration VM problem. As the sake of simplicity, we have simply reused the entropy scheduling code.

DVMS, a distributed approach that dynamically partitions the system and invokes Entropy on each partition.

- Participants: Adrien Lèbre, Flavien Quesnel, Jonathan Pastor, Mario Südholt and Takahiro Hiro-
- Contact: Adrien Lèbre

fuchi

• URL: http://beyondtheclouds.github.io/VMPlaceS/

#### **6.4. ENOS**

Experimental eNvironment for OpenStack

KEYWORDS: OpenStack - Experimentation - Reproducibility FUNCTIONAL DESCRIPTION: Enos workflow :

A typical experiment using Enos is the sequence of several phases:

- enos up : Enos will read the configuration file, get machines from the resource provider and will prepare the next phase - enos os : Enos will deploy OpenStack on the machines. This phase rely highly on Kolla deployment. - enos init-os : Enos will bootstrap the OpenStack installation (default quotas, security rules, ...)
- enos bench : Enos will run a list of benchmarks. Enos support Rally and Shaker benchmarks. - enos backup : Enos will backup metrics gathered, logs and configuration files from the experiment.

- Partner: Orange Labs
- Contact: Adrien Lèbre

## 7. New Results

#### 7.1. Cloud programming and management

#### 7.1.1. Cloud infrastructures

Our contributions regarding cloud infrastructures can be divided into three main topics described below: contributions related to (i) geo-distributed clouds (e.g., Fog and Edge computing), (ii) the convergence of Cloud and HPC infrastructures and (iii) the simulation of virtualized infrastructures.

#### 7.1.1.1. Geo-distributed Clouds

Many academic and industry experts are now advocating a shift from large-centralized Cloud Computing infrastructures to massively small-geo-distributed data centers at the edge of the network. This new paradigm of utility computing is often called Fog and Edge Computing. Advantages of this paradigm are, among others, data-locality that enhances security aspects and response times for latency-critical applications, new energetic options because of reduced size of data centers (e.g., renewable energies), single point of failure avoidance etc. Among the obstacles to the adoption of this model though is the development of a convenient and powerful IaaS system capable of managing a significant number of remote data-centers in a unified way, including monitoring and data management issues in a decentralized environment.

In 2017, we achieved three main contributions toward this challenge.

In [12], we investigate how a holistic monitoring service for a Fog/Edge infrastructure, hosting next generation digital services, should be designed. Although several solutions have been proposed in the past for the monitoring of clusters, grids and cloud systems, none of those is well appropriate to the specific Fog and Edge Computing context. The contributions of this study are: (i) the problem statement, (ii) a classification and a qualitative analysis of major existing solutions, and (iii) a preliminary discussion of the impact of deployment strategies on the monitoring service.

In [6], [39], [17], we present successive studies related to the design and development of a first-class object store service for Fog/Edge facilities. After a deep analysis of major existing solutions (Ceph, Cassandra ...), we designed a proposal that combines Scale-out Network Attached Storage systems (NAS) and IPFS, a BitTorrent-based object store spread throughout the Fog/Edge infrastructure. Without impacting the IPFS advantages particularly in terms of data mobility, the use of a Scale-out NAS on each site reduces the inter-site exchanges that are costly but mandatory for the metadata management in the original IPFS implementation. Several experiments conducted on Grid'5000 testbed are analysed and corroborate, first, the benefit of using an object store service spread at the Edge, and second, the importance of mitigating inter-site accesses. Ongoing activities are related to the management of meta data information in order to benefit from data movements.

Finally, in [26], we introduce the premises of a fog/edge resource management system by leveraging the OpenStack software, a leading IaaS manager in the industry. The novelty of the presented prototype is to operate such an Internet-scale IaaS platform in a fully decentralized manner, using P2P mechanisms to achieve high flexibility and avoid single points of failure. More precisely, we revised the OpenStack Nova service (i.e., virtual machine management and allocation) by leveraging a distributed key/value store instead of the centralized SQL backend. We present experiments that validate the correct behavior and gives performance trends of our prototype through an emulation of several data-centers using Grid'5000 testbed.

#### 7.1.1.2. Cloud and HPC convergence

Geo-distribution of Cloud Infrastructures is not the only current trend of utility computing. Another important challenge is to reach the convergence of Cloud and HPC infrastructures, in other words on-demand HPC. Among challenges of this convergence is, for example, the enhancement of the use of light virtualization techniques on HPC systems, as well as the enhancement of mechanisms to be able to consolidate those VMs without deteriorating the performance of HPC applications, thus minimizing interferences between applications.

In [36], we present Eley, a burst buffer solution that helps to accelerate the performance of Big Data applications while guaranteeing the QoS of HPC applications. To achieve this goal, Eley embraces interference-aware prefetching technique that makes reading data input faster while introducing low interference for HPC applications. Specifically, we equip the prefetcher with five optimization actions including No Action, Full Delay, Partial Delay, Scale Up and Scale Down. It iteratively chooses the best action to optimize the prefetching while guaranteeing the pre-defined QoS requirement of HPC applications (i.e., the deadline constraint for the completion of each I/O phase). Evaluations using a wide range of Big Data and HPC applications show the effectiveness of Eley in reducing the execution time of Big Data applications (shorter map phase) while maintaining the QoS of HPC applications.

#### 7.1.1.3. Virtualization simulation

Finally, it is important to be able to simulate the behavior of proposals for the future architectures. However, current models for virtualized resources are not accurate.

In [32], we present our latest results regarding virtualization abstractions and models for cloud simulation toolkits. Cloud simulators still do not provide accurate models for most Virtual Machine (VM) operations. This leads to incorrect results in evaluating real cloud systems. Following previous works on live-migration, we discuss an experimental study we conducted in order to propose a first-class VM boot time model. Most cloud simulators often ignore the VM boot time or give a naive model to represent it. After studying the relationship between the VM boot time and different system parameters such as CPU utilization, memory usage, I/O and network bandwidth, we introduce a first boot time model that could be integrated into current cloud simulators. Through experiments, we also show that our model correctly reproduced the boot time of a VM under different resources contention.

#### 7.1.2. Deployment and reconfiguration in the Cloud

Being able to manage the new generation of utility computing infrastructures is an important step to build useful system building blocks. The next step is to be able to perform initial deployment of any kind of distributed software (i.e., systems, frameworks or applications) on those infrastructures, thus dealing with a complex process that includes interactions between building blocks such as virtual machine management, optimized deployment plans, monitoring of deployment etc. Such deployment processes cannot be handled manually anymore, for this reason automatic deployments tools have to be designed according to the challenges of new infrastructures (e.g., geo-distribution, hybrid infrastructures etc.). Moreover, as distributed software are more and more dynamic (i.e., reconfiguring themselves at runtime), reconfiguration and self-management capabilities should be handled in an efficient and scalable manner.

#### 7.1.2.1. Initial deployment and placement strategies

When focusing on the initial deployment, many challenges should already need to be addressed such as placement of distributed software onto virtual machines, themselves being placed onto physical resources. This kind of placement problem can be modeled in many different ways, such as linear or constraint programming or graph partitioning. Most of the time a multi-objective NP-hard problem is formulated, and specific heuristics have to be built to reach scalable solutions.

In [18], we present new specific placement constraints and objectives adapted to hybrid clouds infrastructures, and we address this problem through constraint programming. Furthermore we evaluate the expressivity and performance of the solution on a real case study. In the Cloud, if public providers enable simple access to resources for companies and users who have sporadic computation or storage needs, private clouds could sometimes be preferred for security or privacy reasons, or for cost reasons due to a high frequency usage of services. However, in many cases a choice between public or private clouds does not fulfill all requirements of companies and hybrid cloud infrastructures should be preferred. Solutions have already been proposed to address hybrid cloud infrastructures, however most of the time the placement of a distributed software on such infrastructure has to be indicated manually.

In [37], we present a geo-aware graph partitioning method named G-Cut, which aims at minimizing the inter-DC data transfer time of graph processing jobs in geo-distributed DCs while satisfying the WAN usage budget. G-Cut adopts two novel optimization phases which address the two challenges in WAN usage and network heterogeneities separately. G-Cut can be also applied to partition dynamic graphs thanks to its light-weight runtime overhead. We evaluate the effectiveness and efficiency of G-Cut using real-world graphs with both real geo-distributed DCs and simulations. Evaluation results demonstrate that effectiveness of G-Cut in reducing the inter-DC data transfer time and the WAN usage with a low runtime overhead.

Many other challenges than placement rise from the initial deployment. In [20], we present a survey of existing deployment tools that have been used in production to deploy OpenStack, which is a complex distributed system composed of more than a hundred different services. To fully understand how IaaSes are deployed today, we propose in this paper an overall model of the application deployment process that describes each

step with their interactions. This model then serves as the basis to analyse five different deployment tools used to deploy OpenStack in production: Kolla, Enos, Juju, Kubernetes, and TripleO. Finally, a comparison is provided and the results are discussed to extend this analysis.

#### 7.1.2.2. Capacity planning and scheduling

While a placement problem is a discrete problem at a given instant, some other challenges of deployment and reconfiguration may include the time dimension leading to scheduling optimization.

in [30] we have proposed two original workload prediction models for Cloud infrastructures. These two models, respectively based on constraint programming and neural networks, focus on predicting the CPU usage of physical servers in a Cloud data center. The predictions could then be exploited for designing energy-efficient resource allocation mechanisms like scheduling heuristics or over-commitment policies. We also provide an efficient trace generator based on constraint satisfaction problem and using a small amount of real traces. Such a generator can overcome availability issues of extensive real workload traces employed for optimization heuristics validation. While neural networks exhibit higher prediction capabilities, constraint programming techniques are more suitable for trace generation, thus making both techniques complementary.

#### 7.1.2.3. Reconfiguration and self-management

Being able to handle the dynamicity of hardware, system building blocks, middleware and applications is a great challenge of today's and future utility computing systems. On large infrastructures such as Cloud, Fog or Edge Computing, manual administration of such dynamicity is not feasible. The automatic management of reconfiguration, or self-management of software is of great importance to guarantee reliability, fault tolerance, security, and cost and energy optimization.

In [4], in order to improve the self-adaptive behaviors in the context of Component-based Architecture, we design self-adaptive software components based on logical discrete control approaches, in which the self-adaptive behavioural models enrich component controllers with a knowledge not only on events, configurations and past history, but also with possible future configurations. This article provides the description, implementation and discussion of Ctrl-F, a Domain-specific Language whose objective is to provide high-level support for describing these control policies. In [13], we extended Ctrl-F with modularity capabilities. Apart from the benefits of reuse and substitutability of Ctrl-F programs, modularity allows to break down the combinatorial explosion intrinsic to the generation of correct-by-construction controllers in the compilation process of Ctrl-F. A further advantage of modularity is that the executable code, that is, the controllers resulting from that compilation, are loss-coupled and can therefore be deployed and executed in a distributed fashion.

However, higher abstraction-level tools also have to be proposed for reconfiguration. In [21], we introduce ElaScript, a Domain Specific Language (DSL) which offers Cloud administrators a simple and concise way to define complex elasticity-based reconfiguration plans. ElaScript is capable of dealing with both infrastructure and software elasticities, independently or together, in a coordinated way. We validate our approach by first showing the interest to have a DSL offering multiple levels of control for Cloud elasticity, and then by showing its integration with a realistic well-known application benchmark deployed in OpenStack and the Grid'5000 infrastructure testbed.

Finally, self-management can be applied at many different levels of the Cloud paradigm, from infrastructure reconfigurations to application topology reconfigurations. In practice these reconfiguration mechanisms are tightly coupled. For example, a change in the infrastructure could lead to the re-deployment of virtual machines upon it that could lead itself to application reconfigurations. In [27], we advocate that Cloud services, regardless of the layer, may share the same consumer/provider-based abstract model. From that model, we can derive a unique and generic Autonomic Manager (AM) that can be used to manage any XaaS (Everything-as-a-Service) layer defined with that model. The paper proposes such an abstract (although extensible) model along with a generic constraint-based AM that reasons on abstract concepts, service dependencies as well as SLA (Service Level Agreements) constraints in order to find the optimal configuration for the modeled XaaS. The genericity of our approach are shown and discussed through two motivating examples and a qualitative experiment has been carried out in order to show the applicability of our approach as well as to discuss its limitations.

#### 7.2. Energy-aware computing

#### 7.2.1. Renewable energy

In his PhD thesis [1], Md Sabbir Hasan proposes – across three different contributions – how to smartly use green energy at the infrastructure and application levels for further reduction of the corresponding carbon footprints. First, he investigates the options and challenges to integrate different renewable energy sources in a realistic way and proposes a *Cloud energy broker*, which can adjust the availability and price combination to buy Green energy dynamically from the energy market in advance to make a data center partially green. Then, he introduces the concept of *virtualization of green energy*, which can be seen as an alternative to energy storage used in data centers to eliminate the intermittency problem to some extent. With the adoption of this virtualization concept, we can maximize the usage of green energy contrary to energy storage which induces energy losses, while introducing a notion of Green Service Level Agreement based on green energy for service provider and end-users. Finally, he proposes an energy adaptive autoscaling solution to exploit application internals to create green energy awareness in the interactive SaaS applications, while respecting traditional QoS properties.

In [9], we present a scheme for green energy management in the presence of explicit and implicit integration of renewable energy in data center. More specifically we propose three contributions: i) we introduce the concept of *virtualization of green energy* to address the uncertainty of green energy availability, ii) we extend the Cloud Service Level Agreement (CSLA) language <sup>0</sup> to support Green SLA by introducing two new threshold parameters and iii) we introduce green SLA algorithm which leverages the concept of virtualization of green energy to provide per interval specific Green SLA. Experiments were conducted with real workload profile from PlanetLab and server power model from SPECpower to demonstrate that Green SLA can be successfully established and satisfied without incurring higher cost.

In [8], we investigate a thorough analysis of energy consumption and performance trade-off by allowing smart usage of green energy for interactive cloud application. Moreover, we propose an auto-scaler, named as SaaScaler, that implements several control loop based application controllers to satisfy different performance (i.e., response time, availability and user experience) and resource aware metrics (i.e., quality of energy). Based on extensive experiments with RUBiS benchmark and real workload traces using single compute node in Openstack/Grid'5000, results suggest that 13% brown energy consumption can be reduced without deprovisioning any physical or virtual resources at IaaS layer while 29% more users can access the application by dynamically adjusting capacity requirements. In [23], we add to the previous paper the capability of the infrastructure layer to be elastic. We propose a PaaS solution which efficiently utilize the elasticity nature at both infrastructure and application levels, by leveraging adaptation in facing to changing condition i.e., workload burst, performance degradation, quality of energy, etc. While applications are adapted by dynamically re-configuring their service level based on performance and/or green energy availability, the infrastructure takes care of addition/removal of resources based on application's resource demand. Both adaptive behaviors are implemented in separated modules and are coordinated in a sequential manner. We validate our approach by extensive experiments and results obtained over Grid'5000 testbed. Results show that, application can reduce significant amount of brown energy consumption by 35% and daily instance hour cost by 37% compared to a baseline approach.

in [28] we address the problem of improving the utilization of renewable energy for a single data center by using two approaches: opportunistic scheduling and energy storage. Our first result deals with analyzing the workload to find ideal solar panel dimension and battery size, this is used to power the entire workload without any brown energy consumption. However, in reality, either the solar panel dimension or the battery size are limited, and we still have to address the problem of matching the workload consumption and renewable energy production. The second result shows that opportunistic scheduling can reduce the demand for battery size while the renewable energy is sufficient. The last results demonstrate that for different battery sizes and solar panel dimensions, we can find an optimal solution combining both approaches that balances the energy losses due to different causes such as battery efficiency and VM migrations due to consolidation algorithms.

<sup>&</sup>lt;sup>0</sup>http://web.imt-atlantique.fr/x-info/csla

In [5] we presented the EPOC project, focus on energy-aware task execution from the hardware to application's components in the context of a mono-site data center (all resources are in the same physical location) which is connected to the regular electric Grid and to renewable energy sources (such as windmills or solar cells). we have presented the EpoCloud principles, architecture and middleware components. EpoCloud is our prototype, which tackles three major challenges: 1) To optimize the energy consumption of distributed infrastructures and service compositions in the presence of ever more dynamic service applications and ever more stringent availability requirements for services; 2) To design a clever cloud's resource management, which takes advantage of renewable energy availability to perform opportunistic tasks, then exploring the trade-off between energy saving and performance aspects in large-scale distributed system; 3) To investigate energy-aware optical ultra high-speed interconnection networks to exchange large volumes of data (VM memory and storage) over very short periods of time.

in [31] we extend our previous work on PIKA (focus 2 in the EPOC project) and introduced the green energy aware scheduling problem (GEASP) to optimize the energy consumption of a small/medium size data center. Using our model to solve the GEASP, we could optimize the energy consumption of a small/medium size data center in three ways. First, we slightly decrease its overall energy consumption, second we considerably decrease its brown energy consumption and finally we significantly increase its green energy consumption.

#### 7.2.2. Energy-aware consolidation and reconfiguration

In [41] we compared the performance of VMs and containers when consolidating multiple services, in terms of QoS and EE. Our experiments compared two broadly recognized virtualization technologies: KVM for the VM approach, and Docker for the containers. We conclude that Docker outperforms KVM both in QoS and EE. According to our measurements, Docker allows running up to a 21% more services than KVM, when setting a maximum latency of 3,000 ms. In this configuration, Docker offers this service while using a 11.33% less energy than KVM. At a datacenter level, the same computation could run using less servers and less energy per server, accounting for a total of a 28% energy savings inside the datacenter.

The emergence of Internet of Things (IoT) is participating to the increase of data- and energy-hungry applications. As connected devices do not yet offer enough capabilities for sustaining these applications, users perform computation offloading to the cloud. To avoid network bottlenecks and reduce the costs associated to data movement, edge cloud solutions have started being deployed, thus improving the Quality of Service. In [29], we advocated for leveraging on-site renewable energy production in the different edge cloud nodes to green IoT systems while offering improved QoS compared to core cloud solutions. We proposed an analytic model to decide whether to offload computation from the objects to the edge or to the core Cloud, depending on the renewable energy availability and the desired application QoS. This model is validated on our application use-case that deals with video stream analysis from vehicle cameras.

In [33], we address the problem of stragglers (i.e., slow tasks) in Big Data applications. In particular, we introduce a novel straggler detection mechanism to improve the energy efficiency of speculative execution in Hadoop, namely a hierarchical detection mechanism. The goal of this detection mechanism is to identify critical stragglers which strongly affect the job execution times and reduce the number of killed speculative copies which lead to energy waste. We also present an energy-aware copy allocation method to reduce the energy model which expose the trade-off between performance and energy consumption when scheduling a copy. We evaluate our hierarchical detection mechanism and energy-aware copy allocation method on the Grid'5000 testbed using three representative MapReduce applications. Experimental results show a good reduction in the resource wasted on killed speculative copies and an improvement in the energy efficiency compared to state-of-the-art mechanisms.

The increasing size of main memories has lead to the advent of new types of storage systems. These systems propose to keep all data in distributed main memories. In [35], we present a study to characterize the performance and energy consumption of a representative in-memory storage system, namely RAMCloud, to reveal the main factors contributing to performance degradation and energy-inefficiency. Firstly, we reveal that although RAMCloud scales linearly in throughput for read-only applications, it has a non-proportional

power consumption. Mainly because it exhibits the same CPU usage under different levels of access. Secondly, we show that prevalent Web workloads i.e., read-heavy and update-heavy workloads, can impact significantly the performance and the energy consumption. We relate it to the impact of concurrency, i.e., RAMCloud poorly handles its threads under highly-concurrent accesses. Thirdly, we show that replication can be a major bottleneck for performance and energy. Finally, we quantify the overhead of the crash-recovery mechanism in RAMCloud on both energy-consumption and performance.

#### 7.3. Software engineering

#### 7.3.1. Security and privacy

This year, we have developed new results on the security and privacy of cloud systems on all layers of abstraction: a first notion of distributed side-channel attacks on the system-level, privacy-aware middleware storage systems and accountability specifications and implementations on the application level.

#### 7.3.1.1. System-level security for virtualized environments

Isolation on the system-level is a core security challenge for Cloud infrastructures. Similarly, fog and edge infrastructures are based on virtualization to share physical resources among several self-contained execution environments like virtual machines and containers. Yet, isolation may be threatened due to side-channels, created by the virtualization layer or due to the sharing of physical resources like the processor. Side-channel attacks (SCAs) exploit and use such leaky channels to obtain sensitive data. Previous SCAs are local and exploit isolation challenges of virtualized environments to retrieve sensitive information. We have introduced, as a first, the concept of *distributed side-channel attack (DSCA)* that is based on coordinating local attack techniques. We have explored how such attacks can threaten isolation of any virtualized environments such as fog and edge computing. Finally, we have proposed a first set of applicable countermeasures for attack mitigation of DSCAs. [14], [44]

In [24] we presented how the increasing adoption of cloud environments operated with virtualization technology opened the way to a promising hypervisor-based security monitoring approach named Virtual Machine Introspection (VMI). We investigated in Kbin-ID the application of binary code introspection at hypervisor level and analysis mechanisms on all VM kernel binary code, namely all kernel functions, to widely narrow the semantic gap in an automatic and largely OS independent way. Kbin-ID [40] is a novel hypervisor-based main kernel binary code disassembler which enables the hypervisor to locate all VM main kernel binary code and divide it into code blocks given only the address of one arbitrary kernel instruction. In [24] we presented a security use case, we are able to detect running processes that are hidden from Linux task list and ps command output, and more generally that our solution can be used for designing easily automatic and largely kernel portable VMI applications that detect and safely react against malicious activities thanks to the instrumentation of kernel functions.

#### 7.3.1.2. Privacy-Aware Data Storage.

In [34] we propose a cloud storage service that protects the privacy of users by breaking user documents into blocks in order to spread them on several cloud providers. As cloud providers only own a part of the blocks and they do not know the block organization, they can not read user documents. Moreover, the storage service connects directly users and cloud providers without using a third-party as is generally the practice in cloud storage services. Consequently, users do not give critical information (security keys, passwords, etc.) to a third-party.

#### 7.3.1.3. Accountability for Cloud applications.

Nowadays we are witnessing the democratization of cloud services, as a result, more and more end-users (individuals and businesses) are using these services in their daily life. In such scenarios, personal data is generally flowed between several entities. end-users need to be aware of the management, processing, storage and retention of personal data, and to have necessary means to hold service providers accountable for the use of their data. In Walid Benghabrit's thesis we present an accountability framework called Accountability Laboratory (AccLab) that allows to consider accountability from design time to implementation. We developed

a language called Abstract Accountability Language (AAL) that allows to write obligations and accountability policies. This language is based on a formal logic called First Order Linear Temporal Logic (FOTL) which allows to check the consistency of the accountability policies and the compliance between two policies. These policies are translated into a temporal logic called FO-DTL 3, which is associated to a monitoring technique based on formula rewriting. Finally we developed a monitoring tool called Accountability Monitoring (AccMon) which provides means to monitor accountability policies in the context of a real system. These policies are based on FO-DTL 3 logic and the framework can act in both centralized and distributed modes and can run in on-line and off-line modes.

Accountability means to obey a contract and to ensure responsibilities in case of violations. In previous work we defined the Abstract Accountability Language and its AccLab tool support. In order to evaluate the suitability of our language and tool we experiment with the laptop user agreement, one of the policies of the Hope University in Liverpool. While this experiment is still incomplete we are able to draw some preliminary conclusions. The use of FOTL is rather tricky and the only existing prover is not maintained we think to target a first-order logic approach in the future. Natural specifications have traditional issues, for instance missing information, noises, ambiguities etc. But in case of these policies we can say much more. The information system is missing but also most of the details about the auditing process and the rectification aspects (sanction, compensation, explanation, etc). There is also a mixture of proper user behavior with the usage policy which confuses the specifier. A mean to structure the specification is important, we suggest to use templates, and it is also convenient to capture usage and accountability practices.

#### 7.3.2. Software development and programming languages

#### 7.3.2.1. Industrial Internet

In [19], we present a first "vision" paper toward Cloud Manufacturing. More precisely we try to reconsider relationships between Cloud Computing and Cloud Manufacturing based on basic definitions and historical evolution of both worlds. History shows many relations between computer science and manufacturing processes, starting with the initial idea of "digital manufacturing" in the '70s. Since then, advances in computer science have given birth to the *Cloud Computing* (CC) paradigm, where computing resources are seen as a *service* offered to various end-users. Of course, CC has been used as such to improve the IT infrastructure associated to a manufacturing (CMfg) with the perspective of many benefits for both the manufacturing paradigm *Cloud Manufacturing* (CMfg) with the perspective of many benefits for both the manufacturers and their customers. However, despite the usefulness of CC for CMfg, we advocate that considering CC as a core enabling technology for CMfg, as is often put forth in the literature, is limited and should be reconsidered. This paper presents a new core-enabling vision toward CMfg, called *Cloud Anything* (CA). CA is based on the idea of abstracting low-level resources, beyond computing resources, into a set of core control building blocks providing the grounds on top of which any domain could be "cloudified".

#### 7.3.2.2. Cloud and HPC programming

In [43], we deal with testing reproducibility in the context of Cloud elasticity, which requires control of the elasticity behavior, the possibility to select specific resources to be allocated/unallocated, and the coordination of events parallel to the elasticity process. We propose an approach fulfilling those requirements in order to make elasticity testing reproducible. To validate our approach, we perform three experiments on representative bugs on MongoDB and Zookeeper Cloud applications, where our approach succeeds in reproducing all the bugs.

In [7], the Multi-Stencil Framework (MSF) is presented. Even though this framework is applied on HPC numerical simulations, this work can be transposed to many different domains, for instance smart-\* applications of Fog and Edge computing infrastructures, where heterogeneity of computations and programming models have to be handled. As the computation power of modern high performance architectures increases, their heterogeneity and complexity also become more important. One of the big challenges of exascale is to reach programming models that give access to high performance computing (HPC) to many scientists and not only to a few HPC specialists. One relevant solution to ease parallel programming for scientists is Domain Specific Language (DSL). However, one problem to avoid with DSLs is to mutualized existing codes and libraries instead of implementing each solution from scratch. For example, this phenomenon occurs for stencil-based numerical simulations, for which a large number of languages has been proposed without code reuse between them. The Multi-Stencil Framework (MSF) presented in this paper combines a new DSL to component-based programming models to enhance code reuse and separation of concerns in the specific case of stencils. MSF can easily choose one parallelization technique or another, one optimization or another, as well as one back-end implementation or another. It is shown that MSF can reach same performances than a non component-based MPI implementation over 16.384 cores. Finally, the performance model of the framework for hybrid parallelization is validated by evaluations.

## 8. Bilateral Contracts and Grants with Industry

#### 8.1. Bilateral Contracts with Industry

Participants: Adrien Lebre [Contact point], Ronan-Alexandre Cherrueau, Alexandre Van Kempen.

During 2017, we agreed with Orange Labs (Lannion) to conduct a dedicated study on the evaluation of AMQP message bus alternatives within the OpenStack ecosystem. This bilateral contract ("Contrat de Recherche Externalisé") officially started in Sept 2017 for one year. With the allocated budget (100K), we hired a new research engineer, Alexandre Van Kempen. Alexandre Van Kempen works with Ronan-Alexandre Cherrueau (Temporary Resarch Engineer, hired in the context of the MERCURY InriHub) and Matthieu Simonin (Permanent Research Engineer from the Rennes Bretagne Atlantique Center) on conducting this analysis. In addition to extending the EnOS framework previously presented, they are performing several experiments with the support of the OpenStack open-source community (in particulat RedHat). The goal of the study is to identify major drawbacks of the default RabbitMQ solution with respect to the Fog/Edge requirements and evaluate whether some alternatives are available in the open-source ecosytem.

## 9. Partnerships and Cooperations

#### 9.1. Regional Initiatives

#### 9.1.1. RFI Atlanstic 2020

#### 9.1.1.1. CoMe4ACloud

Participants: Thomas Ledoux [coordinator], Frederico Alvares de Oliveira Junior, Zakarea Al Shara.

The high-level objective of the 1-year CoMe4ACloud (Constraints and Model Engineering for Autonomic Clouds) project is to provide an end-to-end solution for autonomic Cloud services. To that end, we rely on techniques of Constraint Programming so as a decision-making tool and Model-driven Engineering to ease the automatic generation of the so-called autonomic managers as well as their synchronization with the managed system (i.e., the Cloud layers).

This year, we got the best paper award of CLOSER 2017 (the 7th International Conference on Cloud Computing and Services Science) [27]. We have also submitted two publications and provided two video-demonstrations of the early results.

CoMe4ACloud is an Atlanstic2020 funded project and supports a post-doc position. The project is led by Ascola research team and involves also AtlanModels and TASC, all of them from the LS2N and situated at IMT Atlantique. See https://come4acloud.github.io for more information.

#### 9.1.1.2. SyMeTRIC

Participant: Jean-Marc Menaud [coordinator].

SyMeTRIC is a regional federated project in Systems Medicine funded by the Pays de la Loire french region. Systems Medicine approaches can be compared to Systems Biology. They aim at integrating several information sources to design and validate bio-models and biomarkers to anticipate and enhance patients follow-up (diagnosis, treatment response prediction, prognosis).

#### 9.2. National Initiatives

#### 9.2.1. CominLabs laboratory of excellence

#### 9.2.1.1. EPOC

Participants: Jean-Marc Menaud [coordinator], Thomas Ledoux, Md Sabbir Hasan, Yunbo Li.

The project EPOC (Energy Proportional and Opportunistic Computing system) is a project running for 4 years. Four other partners collaborate within the project that is coordinated by ASCOLA: Myriads team, and the three institutions ENIB, ENSTB and University of Nantes. In this project, the partners focus on energy-aware task execution from the hardware to application components in the context of a *mono-site* data center (all resources are in the same physical location) which is connected to the *regular electric Grid and to renewable energy sources* (such as windmills or solar cells). Three major challenges are addressed in this context: optimize the energy consumption of distributed infrastructures and service compositions in the presence of ever more dynamic service applications and ever more stringent availability requirements for services; design a clever cloud's resource management which takes advantage of renewable energy availability to perform opportunistic tasks, then exploring the trade-off between energy saving and performance aspects in large-scale distributed system; investigate energy-aware optical ultra high-speed interconnection networks to exchange large volumes of data (VM memory and storage) over very short periods of time.

One of the strengths of the project is to provide a systematic approach, and use a single model for the system (from hard to soft) by mixing constraint programming and behavioral models to manage energy consumption in data centers.

#### 9.2.1.2. PrivGen

Participants: Fatima-Zahra Boujdad, Mario Südholt [coordinator].

PrivGen ("Privacy-preserving sharing and processing of genetic data") is a three-year project that has been started in Oct. 2016 and is conducted by three partners: a team of computer scientists from the LATIM Inserm institute in Brest mainly working on data watermarking techniques, a team of geneticians from an Inserm institute in Rennes working on the gathering and interpretation of genetic data, and the Ascola team. The project provides funding of 330 KEUR altogether with an Ascola share of 120 KEUR.

The project considers challenges related to the outsourcing of genetic data that is in the Cloud by different stakeholders (researchers, organizations, providers, etc.). It tackles several limitations of current security solutions in the cloud, notably the lack of support for different security and privacy properties at once and computations executed at different sites that are executed on behalf of multiple stakeholders.

The partners are working on three main challenges:

- Mechanisms for a continuous digital content protection
- Composition of security and privacy-protection mechanisms
- Distributed processing and sharing of genetic data

The Ascola team is mainly involved in providing solutions for the second and third challenges.

#### 9.2.2. ANR

#### 9.2.2.1. GRECO (ANR)

Participant: Adrien Lebre [Contact point].

The GRECO project (Resource manager for cloud of Things) is an ANR project (ANR-16-CE25-0016) running for 42 months (starting in January 2017 with an allocated budget of 522KEuros, 90KEuro for ASCOLA).

The consortium is composed of 4 partners: Qarnot Computing (coordinator) and 3 academic research group (DATAMOVE and AMA from the LIG in Grenoble and ASCOLA from Inria Rennes Bretagne Atlantique).

The goal of the GRECO project (https://anr-greco.net) is to design a manager for cloud of things. The manager should act at the IaaS, PaaS and SaaS layer of the cloud. One of the principal challenges will consist in handling the execution context of the environment in which the cloud of things operates. Indeed, unlike classical resource managers, connected devices imply to consider new types of networks, execution supports, sensors and new constraints like human interactions. The great mobility and variability of these contexts complexify the modelling of the quality of service. To face this challenge, we intend to innovate in designing scheduling and data management systems that will use machine learning techniques to automatically adapt their behaviour to the execution context. Adaptation here requires a modelling of the recurrent cloud of things usages, the modelling of the dynamics of physical cloud architecture.

#### 9.2.2.2. KerStream (ANR)

Participant: Shadi Ibrahim [Coordinator].

The KerStream project (Big Data Processing: Beyond Hadoop!) is an ANR JCJC (Young Researcher) project (ANR-16-CE25-0014-1) running for 48 months (starting in January 2017 with an allocated budget of 238KEuros).

The goal of the KerStream project is to address the limitations of Hadoop when running Big Data stream applications on large-scale clouds and do a step beyond Hadoop by proposing a new approach, called KerStream, for scalable and resilient Big Data stream processing on clouds. The KerStream project can be seen as the first step towards developing the first French middleware that handles Stream Data processing at Scale.

#### 9.2.3. FSN

#### 9.2.3.1. Hosanna (FSN)

Participants: Jean-Marc Menaud [coordinator], Remy Pottier.

The Hosanna project aims to scientifically and technically addresses the problem of deploying applications on a distributed multi-cloud virtual infrastructure (private cloud, Amazon, OVH, CloudWatt, Numergy etc.) This recent need is an important topic issue highlighted by recent major Outages in 2013 by the biggest players in the cloud such as Amazon or Netflix. This project aims to provide services that allow users to deploy their cloud multi-tier applications on hybrid Clouds infrastructures without any separation between IaaS. The Ascola team is extending its optimization solution to address the task placement problem in a multi-cloud environment and will develop a case study on a secure distributed file system. The project started in 2015 for a duration of 2 years.

#### 9.2.3.2. Hydda (FSN)

Participants: Jean-Marc Menaud [coordinator], Hélène Coullon.

The HYDDA project aims to develop a software solution allowing the deployment of Big Data applications (with hybrid design (HPC/CLoud)) on heterogeneous platforms (cluster, Grid, private Cloud) and orchestrators (Task scheduler like Slurm, Virtual orchestrator (like Nova for OpenStack or Swarm for Docker). The main challenges addressed by the project are: how to propose an easy-to-use service to host (from deployment to elimination) application components that are both typed Cloud and HPC? How propose a service that unifies the HPCaaS (HPC as a service) and the Infrastructure as a Service (IaaS) in order to offer resources on demand and to take into account the specificities of scientific applications? How optimize resources usage of these platforms (CPU, RAM, Disk, Energy, etc.) in order to propose solutions at the least cost?

#### 9.2.4. CPER

#### 9.2.4.1. SeDuCe

Participants: Jean-Marc Menaud [coordinator], Adrien Lebre.

The SeDuCe project (Sustainable Data Centers: Bring Sun, Wind and Cloud Back Together), aims to design an experimental infrastructure dedicated to the study of data centers with low energy footprint. This innovative data center will be the first experimental data center in the world for studying the energy impact of cloud computing and the contribution of renewable energy (solar panels, wind turbines) from the scientific, technological and economic viewpoints. This project is integrated in the national context of grid computing (Grid'5000), and the Constellation project, which will be an inter-node (Pays de la Loire, Brittany).

#### 9.2.5. Inria Project Labs

#### 9.2.5.1. DISCOVERY

**Participants:** Hélène Coullon, Shadi Ibrahim, Adrien Lebre [coordinator], Dimitri Pertin, Ronan-Alexandre Cherrueau, Alexandre Van Kempen, Mario Südholt.

To accommodate the ever-increasing demand for Utility Computing (UC) resources, while taking into account both energy and economical issues, the current trend consists in building larger and larger Data Centers in a few strategic locations. Although such an approach enables UC providers to cope with the actual demand while continuing to operate UC resources through centralized software system, it is far from delivering sustainable and efficient UC infrastructures for future needs.

The DISCOVERY initiative [26] aims at exploring a new way of operating Utility Computing (UC) resources by leveraging any facilities available through the Internet in order to deliver widely distributed platforms that can better match the geographical dispersal of users as well as the ever increasing demand. Critical to the emergence of such locality-based UC (also referred as Fog/Edge Computing) platforms is the availability of appropriate operating mechanisms. The main objective of DISCOVERY is to design, implement, demonstrate and promote a new kind of Cloud Operting System (OS) that will enable the management of such a large-scale and widely distributed infrastructure in an unified and friendly manner.

The consortium is composed of experts in the following research areas: large-scale infrastructure management systems, networking and P2P algorithms. Moreover, two key network operators, namely Orange and RENATER, are involved in the project.

By deploying and using a Fog/Edge OS on backbones, our ultimate vision is to enable large parts of the Internet to be hosted and operated by its internal structure itself: a scalable set of resources delivered by any computing facilities forming the Internet, starting from the larger hubs operated by ISPs, governments and academic institutions, to any idle resources that may be provided by end users.

ASCOLA leads the DISCOVERY IPL and contributes mainly around two axes: VM life cycle management and security concerns.

#### 9.2.6. InriaHub

#### 9.2.6.1. MERCURY

Participants: Ronan-Alexandre Cherrueau, Adrien Lebre [coordinator].

ASCOLA, in particular within the framework of the DISCOVERY initiative has been working on the massively distributed use case since 2013. With the development of several proof-of-concepts around OpenStack, the team has had the opportunity to start an InriaHub action. Named MERCURY, the goal of this action is twofold: (i) support the research development made within the context of DISCOVERY and (ii) favor the transfer toward the OpenStack community.

Further information available at: http://beyondtheClouds.github.io.

#### 9.2.7. Fond d'amorçage IMT Industrie du Futur 2017

#### 9.2.7.1. aLIFE

Participants: Hélène Coullon [coordinator], Jacques Noyé.

The French engineering school IMT Atlantique is organizing the aLIFE workshop between industry and academia, in Nantes during two days on January, 30-31 2018. The objective of this workshop is to share various experiences and success stories, as well as open challenges related to the contribution of software-related research to Factories of the Future, in French apport de l'industrie du Logiciel à l'Industrie du Futur Européenne (aLIFE). To this end, big multinational companies, as well as SMEs and academics will exchange through plenary sessions and discussion panels.

#### 9.2.8. Connect Talent

#### 9.2.8.1. Apollo (Connect Talent)

Participant: Shadi Ibrahim [Coordinator].

The Apollo project (Fast, efficient and privacy-aware Workflow executions in massively distributed Datacenters) is an individual research project "Connect Talent" running for 36 months (starting in November 2017 with an allocated budget of 201KEuros).

The goal of the Apollo project is to investigate novel scheduling policies and mechanisms for fast, efficient and privacy-aware data-intensive workflow executions in massively distributed data-centers.

#### 9.3. European Initiatives

#### 9.3.1. FP7 & H2020 Projects

9.3.1.1. CoqHoTT

Title: Coq for Homotopy Type Theory Programm: H2020 Type: ERC Duration: June 2015 - May 2020

Coordinator: Inria

Inria contact: Nicolas TABAREAU

Every year, software bugs cost hundreds of millions of euros to companies and administrations. Hence, software quality is a prevalent notion and interactive theorem provers based on type theory have shown their efficiency to prove correctness of important pieces of software like the C compiler of the CompCert project. One main interest of such theorem provers is the ability to extract directly the code from the proof. Unfortunately, their democratization suffers from a major drawback, the mismatch between equality in mathematics and in type theory. Thus, significant Coq developments have only been done by virtuosos playing with advanced concepts of computer science and mathematics. Recently, an extension of type theory with homotopical concepts such as univalence is gaining traction because it allows for the first time to marry together expected principles of equality. But the univalence principle has been treated so far as a new axiom which breaks one fundamental property of mechanized proofs: the ability to compute with programs that make use of this axiom. The main goal of the CoqHoTT project is to provide a new generation of proof assistants with a computational version of univalence and use them as a base to implement effective logical model transformation so that the power of the internal logic of the proof assistant needed to prove the correctness of a program can be decided and changed at compile time—according to a trade-off between efficiency and logical expressivity. Our approach is based on a radically new compilation phase technique into a core type theory to modularize the difficulty of finding a decidable type checking algorithm for homotopy type theory. The impact of the CoqHoTT project will be very strong. Even if Coq is already a success, this project will promote it as a major proof assistant, for both computer scientists and mathematicians. CoqHoTT will become an essential tool for program certification and formalization of mathematics.

#### 9.3.1.2. BigStorage

Title: BigStorage: Storage-based Convergence between HPC and Cloud to handle Big Data

Programm: H2020

Duration: January 2015 - December 2018

Coordinator: Universidad politecnica de Madrid

Partners:

Barcelona Supercomputing Center - Centro Nacional de Supercomputacion (Spain)

Ca Technologies Development Spain (Spain)

Commissariat A L Energie Atomique et Aux Energies Alternatives (France)

Deutsches Klimarechenzentrum (Germany)

Foundation for Research and Technology Hellas (Greece)

Fujitsu Technology Solutions (Germany)

Johannes Gutenberg Universitaet Mainz (Germany)

Universidad Politecnica de Madrid (Spain)

Seagate Systems Uk (United Kingdom)

Inria contact: G. Antoniu & A. Lebre

The consortium of this European Training Network (ETN) 'BigStorage: Storage-based Convergence between HPC and Cloud to handle Big Data' will train future data scientists in order to enable them and us to apply holistic and interdisciplinary approaches for taking advantage of a dataoverwhelmed world, which requires HPC and Cloud infrastructures with a redefinition of storage architectures underpinning them - focusing on meeting highly ambitious performance and energy usage objectives. There has been an explosion of digital data, which is changing our knowledge about the world. This huge data collection, which cannot be managed by current data management systems, is known as Big Data. Techniques to address it are gradually combining with what has been traditionally known as High Performance Computing. Therefore, this ETN will focus on the convergence of Big Data, HPC, and Cloud data storage, ist management and analysis. To gain value from Big Data it must be addressed from many different angles: (i) applications, which can exploit this data, (ii) middleware, operating in the cloud and HPC environments, and (iii) infrastructure, which provides the Storage, and Computing capable of handling it. Big Data can only be effectively exploited if techniques and algorithms are available, which help to understand its content, so that it can be processed by decision-making models. This is the main goal of Data Science. We claim that this ETN project will be the ideal means to educate new researchers on the different facets of Data Science (across storage hardware and software architectures, large-scale distributed systems, data management services, data analysis, machine learning, decision making). Such a multifaceted expertise is mandatory to enable researchers to propose appropriate answers to applications requirements, while leveraging advanced data storage solutions unifying cloud and HPC storage facilities.'

#### 9.4. International Initiatives

#### 9.4.1. Inria International Partners

#### 9.4.1.1. Informal International Partners

National University of Singapore (NUS): We collaborate on resource management for workflows in the cloud and optimizing graph processing in geo-distributed data-centers.

#### 9.5. International Research Visitors

#### 9.5.1. Visits to International Teams

#### 9.5.1.1. Research Stays Abroad

HUST and ShenZhen University, China: From October 28 to November 11, S. Ibrahim visited the Services Computing Technology and System Lab at Huazhong university of Science and Technology and the National High Performance Computing Center at Shenzhen University.

## **10.** Dissemination

#### **10.1. Promoting Scientific Activities**

#### 10.1.1. Scientific Events Organisation

- 10.1.1.1. Member of the Organizing Committees
  - J. Noyé has co-organized with M. Aksit the symposium Modularity 2017, colocated with <Programming> 2017 (Brussels).
  - A. Lebre has been Publicity Chair of the Big Graph Processing workshop (co-located with ICDCS'17).
  - A. Lebre has co-organized the CNRS Rescom Summer School 2017 edition (70 persons).

#### 10.1.2. Scientific Events Selection

- 10.1.2.1. Chair of Conference Program Committees
  - A. Lebre and A. Simonet were co-program chair of ICFEC'17.
  - S. Ibrahim was program co-chair of the 17th International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP-2017), Helsinki, Finland, August, 2017.
  - S. Ibrahim was program co-chair of the1st Workshop on the Integration of Extreme Scale Computing and Big Data Management and Analytics (EBDMA 2017), co-located with CCGrid'17, Madrid, Spain, May 2017.

#### 10.1.2.2. Member of the Conference Program Committees

- H. Coullon was member of the program committees of the following conferences: CloudCOM'17, ICFEC'17, ICCS'17, SAC PAPP'17, Compas'17
- S. Ibrahim was member of the program committees of SC'17, Cluster'17, HiPC'17, CCGrid'17, ISPA'17, I-SPAN'17, CloudCom'17, FCST'17, PDSW-DISCS@SC'17, NetBOS@ICNP'17, HPBDC@IPDPS'17, SCRAMBL@CCGrid'17.
- A. Lebre was member of the program committees of HPDC'17, SC'17, CC-GRID'17,CloudCom'2017, and NoF'17.
- T. Ledoux was member of the program committees of the following workshops: ARM'17@Middleware, CrossCloud'17@EuroSys
- J.-M. Menaud was member of the program committees of SDS'17, Xgreen2017, AICT'17, CEIS'17, EEEP'17, Energy'17, SMARTGREENS'17
- J-C. Royer was member of the program committees of CAReMAS, ICIS'2017, SCAI'2017 and WETICE'2017.
- M. Südholt was a member of the program committees of CloudCom'17, ProWeb'17, and Programming'17
- C. Zhou was member of the program committees of FCST'17, ICA3PP'17, EBDMA'@CCGrid'17.

#### 10.1.3. Journal

- 10.1.3.1. Member of the Editorial Boards
  - A. Lebre is an Associate Editor of the IEEE Transactions on BigDat
  - S. Ibrahim is a Guest Editor of IEEE Transactions on Big Data Special Issue on the Integration of Extreme Scale Computing and Big Data Management and Analytics. a Journal.
  - M. Südholt is an Associate Editor of the journals Programming and Modularity (Springer).

10.1.3.2. Reviewer - Reviewing Activities

- H. Coullon has been a reviewer for the following journals: Annals of Telecommunication, Future Generation Computer Systems, Transactions on BigData.
- H. Coullon has been a reviewer for the following conferences: Cluster'17, HPDC'17.
- A. Lebre has been reviewer for the following journals: IEEE Transactions on Network and Service Management, and Journal of Parallel and Distributed Computing.
- A. Lebre has been reviewer for Europar 2017.
- T. Ledoux has been a reviewer for the journal IEEE Transactions on Services Computing.
- J. Noyé has been a reviewer for the Journal of Object Technology.
- S. Ibrahim has been a reviewer for the following journals: IEEE Transactions on Parallel and Distributed Systems, and IEEE Transactions on Big Data.

#### 10.1.4. Invited Talks

- S. Ibrahim has been invited to present a talk at the ResCom summer school (Le Croisic) : "Big Data Processing in the Cloud: Hadoop and Beyond".
- S. Ibrahim has been invited to present a talk at CGCL (Huazhong University of Science and Technology, China, 02/11/2017): "Scalable Big Data Management on clouds and HPC systems".
- S. Ibrahim has been invited to present a talk at ShenZhen University (China, 09/11/2017): "Scalable Big Data Management on clouds and HPC systems".
- A. Lebre has been invited to present a talk at the 11th edition of the CloudControl Workshop serie (Sweden) : "Enos: a Holistic Framework for Conducting Scientific Evaluations of OpenStack".
- A. Lebre has been invited to present a talk at the ResCom summer school (Le Croisic) : "Utility Computing: From Mainframes to Clouds and Beyond!".
- T. Ledoux has been invited for a talk about *frugal* Cloud by the GDS EcoInfo (CNRS) (Grenoble, France, 02/03/2017)
- T. Ledoux has been invited to present the CoMe4ACloud project at the 4th Grenoble Workshop on Autonomic Computing and Control (Grenoble, France, 10/23/2017)
- M. Südholt has been invited to present a talk at the IMT Cybersecurity day on "Privacy and sharing of genomic data."

#### 10.1.5. Scientific Expertise

- S. Ibrahim is Leading the *Resource Management and Scheduling for Data-Intensive HPC Workflows* activity within the JLESC, Joint Inria-Illinois-ANL-BSC-JSC-RIKEN/AICS Laboratory for Extreme-Scale Computing.
- S. Ibrahim is member of Grid'5000 Sites Committee Responsible for the Rennes site.
- A. Lebre is member of the executive committee of the GDR CNRS RSD "Réseau et Système distribué" and Co-leader of the transversal action Virtualization and Clouds of this GDR since 2015.
- A. Lebre is leading the OpenStack "Fog/Edge/Massively Distributed Clouds" Special Interest Group (further information at: https://wiki.openstack.org/wiki/Massively\_Distributed\_Clouds).
- A. Lebre is member of the executive and architect committees of the Grid'5000 GIS (Groupement d'intérêt scientifique).
- J.-M. Menaud is the organizer of "Pôle Science du Logiciel et des Systèmes Distribués" in Laboratoire des Sciences du Numérique à Nantes (LS2N) since June 2015.
- J.-M. Menaud member of the thesis committee Gilles Kahn Award, sponsored by the French Academy of Sciences awarded by the SiF

#### 10.1.6. Research Administration

J. Noyé is deputy head of the Automation, Production and Computer Sciences department of IMT Atlantique.

### **10.2. Teaching - Supervision - Juries**

#### 10.2.1. Supervision

- PhD: Md Sabbir Hasan, "Smart management of renewable energy in Clouds: from infrastructure to application", INSA Rennes, 05 March 2017, advisor: T. Ledoux
- HdR : Adrien Lebre, "Contributions to Large Scale Distributed Systems: The infrastructure view-point", University of Nantes, 1 Sept 2017.
- PhD: Emile Cadorel, director: J-M. Menaud, advisor: H. Coullon
- PhD: Fatima-zahra Boujdad, advisor: Mario Südholt
- PhD: Maverick Chardet, advisors: H. Coullon and A. Lebre
- PhD: Yewan Wang, director: J-M. Menaud.
- PhD: Mohammad Mahdi Bazm: codirectors: Mario Sudholt, J-M. Menaud.
- Postdoc: Dimitri Pertin, advisor: H. Coullon
- Postdoc: Chi Zhou, advisor: S. Ibrahim

#### 10.2.2. Juries

- S. Ibrahim was member of the PhD Committee of Thomas Lambert, "Etude de l'effet de la replication de fichiers d'entree sur l'efficacite et la robustesse d'un ensemble de calcul", University of Bordeaux, September 2017.
- A. Lebre was member of the PhD Committee of Ismael Cuadrado-Cordero, "Microclouds: An Approach for a Network-Aware Energy-Efficient Decentralized Cloud", University of Rennes 1, Feb 2017.
- A. Lebre was member of the PhD Committee of Luis Pineda, "Efficient Support fo Data-Intensive Scientific Workflows on Geo-Distributed Clouds", University of Rennes 1, May 2017.
- A. Lebre was member of the PhD Committee of Aymen Jlassi, "Optimisation de la gestion des ressources sur une plateforme inforamtique de type Big-data basée sur le logiciel Hadoop", University of Tours, Dec 2017.
- T. Ledoux was a member of the PhD committee of Xuan Sang Le, "Software/FPGA Co-design for Edge-computing: Promoting Object-oriented Design", Univ. Bretagne Occidentale, May 2017
- J.-M. Menaud was reviewer of the PhD of Ines de Courchelle, "Vers une meilleure utilisation des énergies renouvelables : application à des bâtiments scientifiques",(Nov. 20, 2017), Toulouse ; Boris Teabe (Oct 12, 2017) "Performance et qualité de service de l'ordonnanceur dans un environnement virtualisé", Toulouse ; Alexis Martin (Jan. 13, 2017) "Infrastructure pour la gestion générique et optimisée des traces d'exécution pour les systèmes embarqués", Grenoble.
- J-C. Royer was member of the PhD comittee of Walid Benghabrit, "A Formal Model for Accountability", IMT Atlantique, October 27 2017 and Gwendal Daniet, "Efficient Persistence, Query and Tranformation of Large Models", IMT Atlantique, November 14, 2017.
- M. Südholt was a member of the PhD committee of Pauline Bolignano, "Formal Models and Verification of Memory Management in a Hypervisor", Inria, May 2017.

## **11. Bibliography**

#### **Publications of the year**

#### **Doctoral Dissertations and Habilitation Theses**

[1] M. S. HASAN.*Smart management of renewable energy in clouds : from infrastructure to application*, INSA de Rennes, May 2017, https://tel.archives-ouvertes.fr/tel-01617963.

- [2] A. LEBRE.Contributions to Large-scale Distributed Systems The infrastructure Viewpoint Toward Fog and Edge Computing as The Next Utility Computing Paradigm?, Université de Nantes, September 2017, Habilitation à diriger des recherches, https://tel.archives-ouvertes.fr/tel-01664343.
- [3] Y. LI.*Resource allocation in a Cloud partially powered by renewable energy sources*, Ecole nationale supérieure Mines-Télécom Atlantique, June 2017, https://tel.archives-ouvertes.fr/tel-01595953.

#### **Articles in International Peer-Reviewed Journal**

- [4] F. ALVARES, E. RUTTEN, L. SEINTURIER. A Domain-specific Language for The Control of Self-adaptive Component-based Architecture, in "Journal of Systems and Software", January 2017, https://hal.archivesouvertes.fr/hal-01450517.
- [5] N. BELDICEANU, B. DUMAS FERIS, P. GRAVEY, S. HASAN, C. JARD, T. LEDOUX, Y. LI, D. LIME, G. MADI-WAMBA, J.-M. MENAUD, P. MOREL, M. MORVAN, M.-L. MOULINARD, A.-C. ORGERIE, J.-L. PAZAT, O. H. ROUX, A. SHARAIHA.*Towards energy-proportional Clouds partially powered by renewable energy*, in "Computing", January 2017, vol. 99, n<sup>o</sup> 1, 20 [DOI: 10.1007/s00607-016-0503-z], https://hal. inria.fr/hal-01340318.
- [6] B. CONFAIS, A. LÈBRE, B. PARREIN. Performance Analysis of Object Store Systems in a Fog and Edge Computing Infrastructure, in "Transactions on Large-Scale Data- and Knowledge-Centered Systems", August 2017 [DOI: 10.1007/978-3-662-55696-2\_2], https://hal.archives-ouvertes.fr/hal-01587459.
- [7] H. COULLON, J. BIGOT, C. PÉREZ. Extensibility and Composability of a Multi-Stencil Domain Specific Framework, in "International Journal of Parallel Programming", November 2017 [DOI: 10.1007/s10766-017-0539-5], https://hal.archives-ouvertes.fr/hal-01650998.
- [8] M. S. HASAN, F. ALVARES, T. LEDOUX, J.-L. PAZAT. Investigating Energy Consumption and Performance Trade-off for Interactive Cloud Application, in "IEEE Transactions on Sustainable Computing", June 2017 [DOI: 10.1109/TSUSC.2017.2714959], https://hal.archives-ouvertes.fr/hal-01540159.
- [9] M. S. HASAN, Y. KOUKI, T. LEDOUX, J.-L. PAZAT. Exploiting Renewable sources: when Green SLA becomes a possible reality in Cloud computing, in "IEEE Transactions on Cloud Computing", June 2017, vol. 5, n<sup>o</sup> 2, p. 249-262 [DOI : 10.1109/TCC.2015.2459710], https://hal.archives-ouvertes.fr/hal-01187907.
- [10] Y. LI, A.-C. ORGERIE, I. RODERO, B. LEMMA AMERSHO, M. PARASHAR, J.-M. MENAUD.End-toend Energy Models for Edge Cloud-based IoT Platforms: Application to Data Stream Analysis in IoT, in "Future Generation Computer Systems", 2017, p. 1-12 [DOI: 10.1016/J.FUTURE.2017.12.048], https:// hal.archives-ouvertes.fr/hal-01673501.
- [11] S. MARTINEZ, M. TISI, R. DOUENCE. Reactive Model Transformation with ATL, in "Science of Computer Programming", March 2017, vol. 136, p. 1 - 16 [DOI: 10.1016/J.SCICO.2016.08.006], https://hal.inria.fr/ hal-01627991.

#### **International Conferences with Proceedings**

[12] M. ABDERRAHIM, M. OUZZIF, K. GUILLOUARD, J. FRANCOIS, A. LÈBRE. A Holistic Monitoring Service for Fog/Edge Infrastructures: a Foresight Study, in "The IEEE 5th International Conference on Future Internet of Things and Cloud (FiCloud 2017)", Prague, Czech Republic, August 2017, p. 337-344 [DOI: 10.1109/FICLOUD.2017.30], https://hal.archives-ouvertes.fr/hal-01591161.

- [13] F. ALVARES, G. DELAVAL, E. RUTTEN, L. SEINTURIER.Language Support for Modular Autonomic Managers in Reconfigurable Software Components, in "SeAC 2017 - 2nd Workshop on Self-Aware Computing @ ICAC'17", Colombus, United States, July 2017, p. 271 - 278 [DOI: 10.1109/ICAC.2017.48], https://hal. inria.fr/hal-01575353.
- [14] M.-M. BAZM, M. LACOSTE, M. SÜDHOLT, J.-M. MENAUD. Side-Channels Beyond the Cloud Edge : New Isolation Threats and Solutions, in "IEEE International Conference on Cyber Security in Networking (CSNet) 2017", Rio de Janeiro, Brazil, 1st Cyber Security in Networking Conference (CSNet'17), October 2017, https://hal.inria.fr/hal-01593144.
- [15] N. BERTHIER, F. ALVARES, H. MARCHAND, G. DELAVAL, E. RUTTEN.Logico-numerical Control for Software Components Reconfiguration, in "CCTA 2017 - IEEE Conference on Control Technology and Applications", Mauna Lani, HI, United States, IEEE, August 2017, p. 1599 - 1606 [DOI: 10.1109/CCTA.2017.8062685], https://hal.inria.fr/hal-01644754.
- [16] R.-A. CHERRUEAU, D. PERTIN, A. SIMONET, A. LEBRE, M. SIMONIN. Toward a Holistic Framework for Conducting Scientific Evaluations of OpenStack, in "The 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)", Madrid, Spain, May 2017, p. 544-548 [DOI: 10.1109/CCGRID.2017.87], https://hal.archives-ouvertes.fr/hal-01664515.
- [17] B. CONFAIS, A. LÈBRE, B. PARREIN. An Object Store Service for a Fog/Edge Computing Infrastructure based on IPFS and Scale-out NAS, in "1st IEEE International Conference on Fog and Edge Computing -ICFEC'2017", Madrid, Spain, 2017 IEEE 1st International Conference on Fog and Edge Computing (ICFEC 2017), May 2017, p. 41-50 [DOI: 10.1109/ICFEC.2017.13], https://hal.archives-ouvertes.fr/hal-01483702.
- [18] H. COULLON, G. LE LOUËT, J.-M. MENAUD. Virtual Machine Placement for Hybrid Cloud using Constraint Programming, in "ICPADS 2017", Shenzhen, China, December 2017, https://hal.archives-ouvertes.fr/hal-01581969.
- [19] H. COULLON, J. NOYÉ. Reconsidering the Relationship between Cloud Computing and Cloud Manufacturing, in "SOHOMA 2017", Nantes, France, October 2017, https://hal.inria.fr/hal-01591113.
- [20] H. COULLON, D. PERTIN, C. PÉREZ. Production Deployment Tools for IaaSes: an Overall Model and Survey, in "The IEEE 5th International Conference on Future Internet of Things and Cloud (FiCloud)", Prague, Czech Republic, August 2017, p. 183-190 [DOI: 10.1109/FICLOUD.2017.51], https://hal.inria.fr/hal-01532489.
- [21] S. DUPONT, S. BOURI, F. ALVARES DE OLIVEIRA, T. LEDOUX.*ElaScript: a DSL for Coding Elasticity in Cloud Computing*, in "32nd ACM Symposium on Applied Computing Track on Cloud Computing", Marrakesh, Morocco, Proceedings of the 32nd ACM Symposium on Applied Computing Track on Cloud Computing, April 2017, https://hal.archives-ouvertes.fr/hal-01400236.
- [22] G. GILBERT. Formalising Real Numbers in Homotopy Type Theory, in "6th ACM SIGPLAN Conference on Certified Programs and Proofs", Paris, France, January 2017, p. 112 - 124, https://arxiv.org/abs/1610.05072 [DOI: 10.1145/3018610.3018614], https://hal.inria.fr/hal-01449326.
- [23] M. S. HASAN, F. ALVARES DE OLIVEIRA JR., T. LEDOUX.GPaaScaler: Green Energy aware Platform Scaler for Interactive Cloud Application, in "IEEE/ACM International Conference on Utility and Cloud Computing", Austin, United States, December 2017, https://hal.archives-ouvertes.fr/hal-01650136.

- [24] Y. HEBBAL, L. SYLVIE, J.-M. MENAUD.*Hidden Process Detection using Kernel Functions Instrumentation*, in "IEEE Conference on Dependable and Secure Computing (DSC)", Taiwan, Taiwan, August 2017, https:// hal.inria.fr/hal-01520864.
- [25] Y. HEBBAL, L. SYLVIE, J.-M. MENAUD.K-binID: Kernel Binary Code Identification for Virtual Machine Introspection, in "IEEE Conference on Dependable and Secure Computing (DSC)", Taiwan, Taiwan, August 2017, https://hal.inria.fr/hal-01520867.
- [26] A. LEBRE, J. PASTOR, A. SIMONET, F. DESPREZ. Revising OpenStack to Operate Fog/Edge Computing infrastructures, in "IEEE International Conference on Cloud Engineering", Vancouver, France, April 2017, https://hal.inria.fr/hal-01273427.
- [27] Best Paper
  - J. LEJEUNE, F. ALVARES, T. LEDOUX. *Towards a generic autonomic model to manage Cloud Services*, in "The 7th International Conference on Cloud Computing and Services Science (CLOSER 2017)", Porto, Portugal, April 2017, https://hal.archives-ouvertes.fr/hal-01511360.
- [28] Y. LI, A.-C. ORGERIE, J.-M. MENAUD.Balancing the use of batteries and opportunistic scheduling policies for maximizing renewable energy consumption in a Cloud data center, in "PDP 2017 - 25th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing", St Petersburg, Russia, March 2017, https://hal.inria.fr/hal-01432752.
- [29] Y. LI, A.-C. ORGERIE, I. RODERO, M. PARASHAR, J.-M. MENAUD. Leveraging Renewable Energy in Edge Clouds for Data Stream Analysis in IoT, in "CCGrid 2017 - IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing", Madrid, Spain, May 2017, https://hal.inria.fr/hal-01472358.
- [30] G. MADI WAMBA, Y. LI, A.-C. ORGERIE, N. BELDICEANU, J.-M. MENAUD. Cloud workload prediction and generation models, in "SBAC-PAD: International Symposium on Computer Architecture and High Performance Computing", Campinas, Brazil, October 2017, https://hal.archives-ouvertes.fr/hal-01578354.
- [31] G. MADI WAMBA, Y. LI, A.-C. ORGERIE, N. BELDICEANU, J.-M. MENAUD.Green energy aware scheduling problem in virtualized datacenters, in "ICPADS 2017", Shenzen, China, December 2017, https:// hal.archives-ouvertes.fr/hal-01582936.
- [32] T. L. NGUYEN, A. LEBRE. Virtual Machine Boot Time Model, in "PDP 2017 25th Euromicro International Conference on Parallel, Distributed and Network-based Processing", St Peterbourg, Russia, March 2017, p. 430 - 437 [DOI: 10.1109/PDP.2017.58], https://hal.inria.fr/hal-01586932.
- [33] T.-D. PHAN, S. IBRAHIM, A. C. ZHOU, G. AUPY, G. ANTONIU. Energy-Driven Straggler Mitigation in MapReduce, in "Euro-Par'17 - 23rd International European Conference on Parallel and Distributed Computing", Santiago de Compostela, Spain, August 2017, https://hal.inria.fr/hal-01560044.
- [34] R. POTTIER, J.-M. MENAUD. Privacy-Aware Data Storage in Cloud Computing, in "International Conference on Cloud Computing and Services Science (CLOSER)", Porto, Portugal, April 2017, 8, https://hal.inria.fr/hal-01472133.

- [35] Y. TALEB, S. IBRAHIM, G. ANTONIU, T. CORTES. Characterizing Performance and Energy-Efficiency of The RAMCloud Storage System, in "The 37th IEEE International Conference on Distributed Computing Systems (ICDCS 2017)", Atlanta, United States, June 2017, https://hal.inria.fr/hal-01496959.
- [36] O. YILDIZ, A. C. ZHOU, S. IBRAHIM.*Eley: On the Effectiveness of Burst Buffers for Big Data Processing in HPC systems*, in "Cluster'17-2017 IEEE International Conference on Cluster Computing", Hawaii, United States, September 2017, https://hal.inria.fr/hal-01570737.
- [37] A. C. ZHOU, S. IBRAHIM, B. HE.On Achieving Efficient Data Transfer for Graph Processing in Geo-Distributed Datacenters, in "ICDCS'17- The 37th IEEE International Conference on Distributed Computing Systems (ICDCS 2017)", Atlanta, United States, June 2017, https://hal.inria.fr/hal-01560187.

#### **National Conferences with Proceeding**

- [38] S. BOULIER, P.-M. PÉDROT, N. TABAREAU.Modèles de la théorie des types donnés par traduction de programme, in "28ièmes Journées Francophones des Langages Applicatifs", Gourette, France, January 2017, https://hal.archives-ouvertes.fr/hal-01503089.
- [39] B. CONFAIS, A. LÈBRE, B. PARREIN. An object store for Fog infrastructures based on IPFS and a Scale-Out NAS, in "RESCOM 2017", Le Croisic, France, RESCOM 2017 Summer school, June 2017, 2, https://hal. archives-ouvertes.fr/hal-01559065.
- [40] Y. HEBBAL, L. SYLVIE, J.-M. MENAUD.K-binID : Identification de code binaire noyau à des fins d'introspection de machine virtuelle, in "Conférence d'informatique en Parallélisme, Architecture et Système (ComPAS)", Sophia Antipolis, France, June 2017, https://hal.inria.fr/hal-01520863.

#### **Conferences without Proceedings**

[41] I. CUADRADO-CORDERO, A.-C. ORGERIE, J.-M. MENAUD.Comparative Experimental Analysis of the Quality-of-Service and Energy-Efficiency of VMs and Containers' Consolidation for Cloud Applications, in "SoftCOM: International Conference on Software, Telecommunications and Computer Networks", Split, Croatia, September 2017, https://hal.archives-ouvertes.fr/hal-01578325.

#### Scientific Books (or Scientific Book chapters)

[42] D. SAUCEZ, A. LÈBRE, S. SECCI (editors). RESCOM 2017 Summer school, CNRS, June 2017, https://hal. inria.fr/hal-01558074.

#### **Scientific Popularization**

[43] M. ALBONICO, J.-M. MOTTU, G. SUNYÉ, F. ALVARES. Making Cloud-based Systems Elasticity Testing Reproducible, in "7th International Conference on Cloud Computing and Services Science", Porto, Portugal, April 2017, https://hal.inria.fr/hal-01471916.

#### **Other Publications**

[44] M.-M. BAZM, M. LACOSTE, M. SÜDHOLT, J.-M. MENAUD.Side Channels in the Cloud: Isolation Challenges, Attacks, and Countermeasures, March 2017, working paper or preprint, https://hal.inria.fr/hal-01591808. [45] S. BOULIER, N. TABAREAU.*Model structure on the universe in a two level type theory*, 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01579822.

#### **References in notes**

- [46] M. AKŞIT, S. CLARKE, T. ELRAD, R. E. FILMAN (editors). *Aspect-Oriented Software Development*, Addison-Wesley Professional, September 2004.
- [47] C. ALLAN, P. AVGUSTINOV, A. S. CHRISTENSEN, L. HENDREN, S. KUZINS, O. LHOTÁK, O. DE MOOR, D. SERENI, G. SITTAMPALAM, J. TIBBLE. Adding trace matching with free variables to AspectJ, in "ACM Conference on Object-Oriented Programming, Systems and Languages (OOPSLA)", R. P. GABRIEL (editor), ACM Press, 2005.
- [48] R. ALLEN, D. GARLAN. A Formal Basis for Architectural Connection, in "ACM Transactions on Software Engineering and Methodology", July 1997, vol. 6, n<sup>o</sup> 3, p. 213–49.
- [49] J. H. ANDREWS. Process-Algebraic Foundations of Aspect-Oriented Programming, in "Proceedings of the 3rd International Conference on Metalevel Architectures and Separation of Crosscutting Concerns", Lecture Notes in Computer Science, 2001, vol. 2192, p. 187–209.
- [50] M. ARMBRUST, A. FOX, R. GRIFFITH, A. D. JOSEPH, R. KATZ, A. KONWINSKI, G. LEE, D. PATTERSON, A. RABKIN, I. STOICA, M. ZAHARIA. *A view of cloud computing*, in "Communications of the ACM", 2010, vol. 53, n<sup>0</sup> 4, p. 50–58.
- [51] T. H. AUSTIN, C. FLANAGAN. Multiple facets for dynamic information flow, in "Proceedings of the 39th annual ACM SIGPLAN-SIGACT symposium on Principles of programming languages", New York, USA, POPL '12, ACM, 2012, p. 165–178, http://doi.acm.org/10.1145/2103656.2103677.
- [52] A. BELOGLAZOV, R. BUYYA. Energy efficient resource management in virtualized cloud data centers, in "in: Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, CCGRID'10, IEEE Computer Society", 2010, p. 826–831.
- [53] L. D. BENAVIDES NAVARRO, M. SÜDHOLT, W. VANDERPERREN, B. DE FRAINE, D. SUVÉE. Explicitly distributed AOP using AWED, in "Aspect-Oriented Software Development (AOSD)", ACM Press, March 2006, p. 51-62.
- [54] G. S. BLAIR, G. COULSON, P. ROBIN, M. PAPATHOMAS. An architecture for next generation middleware, in "Proceedings of the IFIP International Conference on Distributed Systems Platforms and Open Distributed Processing", Springer-Verlag, 1998.
- [55] A. BRACCIALIA, A. BROGI, C. CANAL. A formal approach to component adaptation, in "Journal of Systems and Software", 2005.
- [56] R. E. BROWN, E. R. MASANET, B. NORDMAN, W. F. TSCHUDI, A. SHEHABI, J. STANLEY, J. G. KOOMEY, D. A. SARTOR, P. T. CHAN. Server and Data Center Energy Efficiency: Public Law 109-431, in "Report to Congress", 06/2008 2008.
- [57] S. CAPECCHI, I. CASTELLANI, M. DEZANI-CIANCAGLINI, T. REZK. Session Types for Access and Information Flow Control, in "CONCUR 2010 - Concurrency Theory, 21th International Conference, CONCUR

2010, Paris, France, August 31-September 3, 2010. Proceedings", P. GASTIN, F. LAROUSSINIE (editors), Lecture Notes in Computer Science, Springer, 2010, vol. 6269, p. 237–252, http://dx.doi.org/10.1007/978-3-642-15375-4\_17.

- [58] E. M. CLARKE, O. GRUMBERG, D. A. PELED. Model Checking, The MIT Press, Cambridge, Massachusetts, 1999.
- [59] A. COLYER, A. CLEMENT. Large-scale AOSD for Middleware, in "Proceedings of the 3rd ACM Int. Conf. on Aspect-Oriented Software Development (AOSD), Lancaster", K. LIEBERHERR (editor), ACM Press, 2004, p. 56–65.
- [60] F. DEREMER, H. H. KRON. Programming-in-the-large versus programming-in-the-small, in "IEEE Transactions on Software Engineering", 1976, vol. SE-2, n<sup>o</sup> 2, p. 80-86.
- [61] G. DECKER, O. KOPP, F. LEYMANN, M. WESKE.BPEL4Chor: Extending BPEL for Modeling Choreographies, in "IEEE International Conference on Web Services (ICWS 2007)", IEEE Computer Society, 2007, p. 296–303.
- [62] E. W. DIJKSTRA. On the role of scientific thought, in "Selected Writings on Computing: A Personal Perspective", Springer-Verlag, 1974, p. 60–66, Published in 1982.
- [63] R. DOUENCE, P. FRADET, M. SÜDHOLT. A framework for the detection and resolution of aspect interactions, in "Proceedings of the ACM SIGPLAN/SIGSOFT Conference on Generative Programming and Component Engineering (GPCE'02)", Lecture Notes in Computer Science, Springer-Verlag, October 2002, vol. 2487, p. 173–188, http://hal.inria.fr/inria-00072153.
- [64] R. DOUENCE, P. FRADET, M. SÜDHOLT. *Trace-Based Aspects*, in "Aspect-Oriented Software Development", M. AKŞIT, S. CLARKE, T. ELRAD, R. E. FILMAN (editors), Addison-Wesley, 2004, p. 201-218.
- [65] R. DOUENCE, O. MOTELET, M. SÜDHOLT. *A formal definition of crosscuts*, in "Proceedings of the 3rd International Conference on Metalevel Architectures and Separation of Crosscutting Concerns", Lecture Notes in Computer Science, Springer-Verlag, 2001, vol. 2192, p. 170–186.
- [66] R. DOUENCE, D. LE BOTLAN, J. NOYÉ, M. SÜDHOLT. *Concurrent Aspects*, in "Proc. of the Int. ACM Conf. on Generative Programming and Component Engineering (GPCE)", ACM Press, October 2006, p. 79-88.
- [67] H. FOSTER, S. UCHITEL, J. MAGEE, J. KRAMER.*Model-based Verification of Web Service Compositions*, in "Proceedings of the 18th IEEE Int. Conf. on Automated Software Engineering (ASE'03)", IEEE Computer Society, 2003, p. 152–163.
- [68] I. FOSTER, Y. ZHAO, I. RAICU, S. LU. Cloud computing and grid computing 360-degree compared, in "Grid Computing Environments Workshop, 2008. GCE'08", Ieee, 2008, p. 1–10.
- [69] A. FUGGETTA, G. P. PICCO, G. VIGNA. Understanding Code Mobility, in "IEEE Transactions on Software Engineering", May 1998, vol. 24, n<sup>o</sup> 5, p. 342–361.
- [70] GREENPEACE.*Make IT green: Cloud computing and its contribution to climate change*, Greenpeace International, March 2010.

- [71] K. HONDA, N. YOSHIDA, M. CARBONE. Multiparty asynchronous session types, in "Proceedings of the 35th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2008, San Francisco, California, USA, January 7-12, 2008", G. C. NECULA, P. WADLER (editors), ACM, 2008, p. 273–284, http:// www.doc.ic.ac.uk/~yoshida/multiparty/multiparty.pdf.
- [72] G. KICZALES, E. HILSDALE, J. HUGUNIN, M. KERSTEN, J. PALM, W. G. GRISWOLD. An Overview of AspectJ, in "ECOOP 2001 — Object-Oriented Programming 15th European Conference, Budapest Hungary", Berlin, J. L. KNUDSEN (editor), Lecture Notes in Computer Science, Springer-Verlag, June 2001, vol. 2072, p. 327–353, http://www.eclipse.org/aspectj/.
- [73] G. KICZALES. Aspect Oriented Programming, in "Proc. of the Int. Workshop on Composability Issues in Object-Orientation (CIOO'96) at ECOOP", July 1996, Selected paper published by dpunkt press, Heidelberg, Germany.
- [74] G. KICZALES, J. DES RIVIERES, DANIEL G. BOBROW. *The Art of the Meta-Object Protocol*, MIT Press, Cambridge (MA), USA, 1991.
- [75] J. KIENZLE, R. GUERRAOUI. AOP Does It Make Sense? The Case of Concurrency and Failures, in "16th European Conference on Object-Oriented Programming (ECOOP'2002)", Malaga, Spain, B. MAGNUSSON (editor), Lecture Notes in Computer Science, Springer-Verlag, 2002.
- [76] T. LEDOUX.OpenCorba: a Reflective Open Broker, in "ACM Meta-Level Architectures and Reflection, Second International Conference, Reflection'99", Saint-Malo, France, P. COINTE (editor), Lecture Notes in Computer Science, Springer-Verlag, July 1999, vol. 1616, p. 197–214.
- [77] M. MCILROY.*Mass produced software components*, in "Mass produced software components", Garmish, Germany, P. NAUR, B. RANDELL (editors), NATO Science Committee, October 1968, p. 138-155.
- [78] N. MEDVIDOVIC, R. N. TAYLOR.A Classification and Comparison Framework for Software Architecture Description Languages, in "IEEE Transactions on Software Engineering", January 2000, vol. 26, n<sup>o</sup> 1, p. 70-93.
- [79] M. MERNIK, J. HEERING, A. M. SLOANE. *When and How to Develop Domain-Specific Languages*, in "ACM Computing Surveys", December 2005, vol. 37, n<sup>o</sup> 4, p. 316-344.
- [80] L. MIKHAJLOV, E. SEKERINSKI. *A study of the fragile base class*, in "A study of the fragile base class", Brussels, Belgium, E. JUL (editor), Lecture Notes in Computer Science, July 1998, vol. 1445, p. 355-382.
- [81] D. H. NGUYEN, M. SÜDHOLT. VPA-based aspects: better support for AOP over protocols, in "4th IEEE International Conference on Software Engineering and Formal Methods (SEFM'06)", IEEE Computer Society Press, September 2006.
- [82] O. NIERSTRASZ. Regular Types for Active Objects, in "Object-Oriented Software Composition", O. NIER-STRASZ, D. TSICHRITZIS (editors), Prentice Hall, 1995, chap. 4, p. 99–121.
- [83] M. NISHIZAWA, S. CHIBA, M. TATSUBORI. Remote Pointcut A Language Construct for Distributed AOP, in "Proceedings of the 3rd ACM Int. Conf. on Aspect-Oriented Software Development (AOSD), Lancaster", ACM Press, 2004.

- [84] D. L. PARNAS. On the criteria for decomposing systems into modules, in "Communications of the ACM", December 1972, vol. 15, n<sup>o</sup> 12, p. 1053-1058.
- [85] S. PEARSON. Toward Accountability in the Cloud, in "Internet Computing, IEEE", July-Aug. 2011, vol. 15, n<sup>o</sup> 4, p. 64-69, http://dx.doi.org/10.1109/MIC.2011.98.
- [86] F. PLASIL, S. VISNOVSKY. Behavior Protocols for Software Components, in "Transactions on Software Engineering", January 2002, vol. 28, n<sup>o</sup> 9.
- [87] F. PUNTIGAM. Coordination Requirements Expressed in Types for Active Objects, in "ECOOP'97—Object-Oriented Programming", M. AKŞIT, S. MATSUOKA (editors), Lecture Notes in Computer Science, Springer-Verlag, 1997, vol. 1241, p. 367–388.
- [88] N. SHARMA, S. BARKER, D. IRWIN, P. SHENOY.Blink: managing server clusters on intermittent power, in "SIGARCH Comput. Archit. News", March 2011, vol. 39, p. 185–198, http://dx.doi.org/10.1145/1961295. 1950389.
- [89] M. SHAW, D. GARLAN. Software Architecture: Perspectives on an Emerging Discipline, Prentice-Hall, 1996.
- [90] B. C. SMITH.*Reflection and Semantics in LISP*, Xerox Palto Alto Research Center, Palo Alto, 1984, n<sup>o</sup> P84-00030.
- [91] S. SOARES, E. LAUREANO, P. BORBA.*Implementing distribution and persistence aspects with AspectJ*, in "Proceedings of the 17th ACM conference on Object-oriented programming, systems, languages, and applications (OOPSLA-02)", C. NORRIS, J. J. B. FENWICK (editors), ACM SIGPLAN Notices, ACM Press, November 4–8 2002, vol. 37, 11, p. 174–190.
- [92] S. SUNDARESWARAN. Ensuring Distributed Accountability for Data Sharing in the Cloud, in "Dependable and Secure Computing", 2012, vol. 9, http://ieeexplore.ieee.org/xpls/abs\_all.jsp?arnumber=6165313.
- [93] R. J. WALKER, K. VIGGERS. Implementing Protocols via Declarative Event Patterns, in "Proceedings of the ACM SIGSOFT International Symposium on Foundations of Software Engineering (FSE-12)", ACM Press, 2004, p. 159 - 169.
- [94] M. WAND, G. KICZALES, C. DUTCHYN. A Semantics for Advice and Dynamic Join Points in Aspect-Oriented Programming, in "ACM Transactions on Programming Languages and Systems (TOPLAS)", 2004, vol. 26, n<sup>o</sup> 5, p. 890–910.
- [95] D. M. YELLIN, R. E. STROM. Protocol specifications and component adaptors, in "ACM Transactions of Programming Languages and Systems", March 1997, vol. 19, n<sup>o</sup> 2, p. 292–333.
- [96] Q. ZHANG, L. CHENG, R. BOUTABA. *Cloud computing: state-of-the-art and research challenges*, in "Journal of internet services and applications", 2010, vol. 1, n<sup>o</sup> 1, p. 7–18.
- [97] A. VAN DEURSEN, P. KLINT, J. VISSER. Domain-Specific Languages: An Annotated Bibliography, in "ACM SIGPLAN Notices", June 2000, vol. 35, n<sup>o</sup> 6, p. 26-36.

## **Team ASPI**

# Applications of interacting particle systems to statistics

Inria teams are typically groups of researchers working on the definition of a common project, and objectives, with the goal to arrive at the creation of a project-team. Such project-teams may include other partners (universities or research institutions).

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Stochastic approaches

## **Table of contents**

1.	Personnel	. 73
2.	Overall Objectives	. 74
3.	Research Program	. 75
	3.1. Interacting Monte Carlo methods and particle approximation of Feynman–Kac distributions	75
	3.2. Multilevel splitting for rare event simulation	76
	3.3. Statistical learning: pattern recognition and nonparametric regression	78
4.	Application Domains	. 79
	4.1. Localisation, navigation and tracking	79
	4.2. Rare event simulation	79
5.	New Results	. 80
	5.1. Central limit theorem for adaptive multilevel splitting	80
	5.2. Adaptive multilevel splitting for Monte Carlo particle transport	80
	5.3. Weak overdamped limit theorem for Langevin processes	80
	5.4. Particle–Kalman filter for structural health monitoring	80
	5.5. Reduced modeling of unknown trajectories	81
	5.6. Model reduction from partial observations	81
	5.7. Low–rank dynamic mode decomposition: optimal solution in polynomial time	82
	5.8. Optimal kernel–based dynamic mode decomposition	82
	5.9. Non parametric state–space model for missing–data imputation	82
6.	Bilateral Contracts and Grants with Industry	. 83
	6.1.1. Hybrid indoor navigation — PhD project at CEA LETI	83
-	6.1.2. Bayesian tracking from raw data — CIFRE grant with DCNS Nantes	83
7.	Partnerships and Cooperations	. 84
	7.2. Nectional initiatives	84
	7.2. National initiatives	84
	7.2.1. Computational Statistics and Molecular Simulation (COSMOS) — ANK challeng	3e
	722 Advanged Geophysical Reduced Order Model Construction from Image Observation	04
	(CERONIMO) AND programma Jaunas Charabausas at Jaunas Charabaura	05
	(GEROININO) — AINK programme jeunes Chercheuses et jeunes Chercheurs	0 <i>J</i> 95
	7.3. European initiatives 7.3.1 Molecular Simulation: Modeling Algorithms and Mathematical Analysis (MSMaths) -	- 85
	FRC Consolidator Grant	85
	7.3.2 Design of Desalination Systems Based on Ontimal Usage of Multiple Renewable Energy	w
	Sources (DESIRES) — FRANFTMED NEXUS_14_049	59 86
	7.4 International initiatives	86
	7.5 International research visitors	86
8	Dissemination	86
0.	8.1. Promoting scientific activities	86
	8.1.1. Scientific events organisation	86
	8.1.2. Participation in workshops, seminars, lectures, etc.	87
	8.1.3. Research administration	87
	8.2. Teaching, supervision, thesis committees	87
	8.2.1. Teaching	87
	8.2.2. Supervision	88
	8.2.3. Thesis committees	88
9.	Bibliography	. 89
# **Team ASPI**

*Creation of the Project-Team: 2005 January 10, updated into Team: 2017 January 01, end of the Team: 2017 December 31* 

#### **Keywords:**

#### **Computer Science and Digital Science:**

- A3.4.5. Bayesian methods
- A3.4.7. Kernel methods
- A5.9.2. Estimation, modeling
- A6.1.1. Continuous Modeling (PDE, ODE)
- A6.1.2. Stochastic Modeling (SPDE, SDE)
- A6.2.2. Numerical probability
- A6.2.3. Probabilistic methods
- A6.2.4. Statistical methods
- A6.3.2. Data assimilation
- A6.3.4. Model reduction

#### **Other Research Topics and Application Domains:**

- B2.3. Epidemiology
- B3.2. Climate and meteorology
- B3.3.2. Water: sea & ocean, lake & river
- B3.3.4. Atmosphere
- B7.1.3. Air traffic
- B9.4.3. Physics
- B9.4.4. Chemistry

# 1. Personnel

## **Research Scientists**

François Le Gland [Team leader, Inria, Senior Researcher] Frédéric Cérou [Inria, Researcher] Patrick Héas [Inria, Researcher] Mathias Rousset [Inria, Researcher, HDR]

#### **Faculty Member**

Valérie Monbet [Univ de Rennes I, Professor, HDR]

#### **External Collaborator**

Arnaud Guyader [Univ Pierre et Marie Curie, HDR]

#### **PhD Students**

Thi Tuyet Trang Chau [Univ. Rennes 1] Ramatoulaye Dabo [Univ. Assane Seck de Ziguinchor and Univ. Rennes 1, co–tutelle] Audrey Cuillery [Naval Group] Kersane Zoubert–Ousseni [CEA, until Nov 2017]

#### **Post-Doctoral Fellow**

Hassan Maatouk [Inria, until Aug 2017]

#### Administrative Assistant

Fabienne Cuyollaa [Inria]

# 2. Overall Objectives

# 2.1. Overall Objectives

The scientific objectives of ASPI are the design, analysis and implementation of interacting Monte Carlo methods, also known as particle methods, with focus on

- statistical inference in hidden Markov models and particle filtering,
- risk evaluation and simulation of rare events,
- global optimization.

The whole problematic is multidisciplinary, not only because of the many scientific and engineering areas in which particle methods are used, but also because of the diversity of the scientific communities which have already contributed to establish the foundations of the field

target tracking, interacting particle systems, empirical processes, genetic algorithms (GA), hidden Markov models and nonlinear filtering, Bayesian statistics, Markov chain Monte Carlo (MCMC) methods, etc.

Intuitively speaking, interacting Monte Carlo methods are sequential simulation methods, in which particles

- *explore* the state space by mimicking the evolution of an underlying random process,
- *learn* their environment by evaluating a fitness function,
- and *interact* so that only the most successful particles (in view of the fitness function) are allowed to survive and to get offsprings at the next generation.

The effect of this mutation / selection mechanism is to automatically concentrate particles (i.e. the available computing power) in regions of interest of the state space. In the special case of particle filtering, which has numerous applications under the generic heading of positioning, navigation and tracking, in

target tracking, computer vision, mobile robotics, wireless communications, ubiquitous computing and ambient intelligence, sensor networks, etc.,

each particle represents a possible hidden state, and is replicated or terminated at the next generation on the basis of its consistency with the current observation, as quantified by the likelihood function. With these genetic-type algorithms, it becomes easy to efficiently combine a prior model of displacement with or without constraints, sensor-based measurements, and a base of reference measurements, for example in the form of a digital map (digital elevation map, attenuation map, etc.). In the most general case, particle methods provide approximations of Feynman–Kac distributions, a pathwise generalization of Gibbs–Boltzmann distributions, by means of the weighted empirical probability distribution associated with an interacting particle system, with applications that go far beyond filtering, in

simulation of rare events, global optimization, molecular simulation, etc.

The main applications currently considered are geolocalisation and tracking of mobile terminals, terrain–aided navigation, data fusion for indoor localisation, optimization of sensors location and activation, risk assessment in air traffic management, protection of digital documents.

# 3. Research Program

# 3.1. Interacting Monte Carlo methods and particle approximation of Feynman–Kac distributions

Monte Carlo methods are numerical methods that are widely used in situations where (i) a stochastic (usually Markovian) model is given for some underlying process, and (ii) some quantity of interest should be evaluated, that can be expressed in terms of the expected value of a functional of the process trajectory, which includes as an important special case the probability that a given event has occurred. Numerous examples can be found, e.g. in financial engineering (pricing of options and derivative securities) [36], in performance evaluation of communication networks (probability of buffer overflow), in statistics of hidden Markov models (state estimation, evaluation of contrast and score functions), etc. Very often in practice, no analytical expression is available for the quantity of interest, but it is possible to simulate trajectories of the underlying process. The idea behind Monte Carlo methods is to generate independent trajectories of this process or of an alternate instrumental process, and to build an approximation (estimator) of the quantity of interest in terms of the weighted empirical probability distribution associated with the resulting independent sample. By the law of large numbers, the above estimator converges as the size N of the sample goes to infinity, with rate  $1/\sqrt{N}$  and the asymptotic variance can be estimated using an appropriate central limit theorem. To reduce the variance of the estimator, many variance reduction techniques have been proposed. Still, running independent Monte Carlo simulations can lead to very poor results, because trajectories are generated *blindly*, and only afterwards are the corresponding weights evaluated. Some of the weights can happen to be negligible, in which case the corresponding trajectories are not going to contribute to the estimator, i.e. computing power has been wasted.

A major breakthrough made in the mid 90's, has been the introduction of interacting Monte Carlo methods, also known as sequential Monte Carlo (SMC) methods, in which a whole (possibly weighted) sample, called *system of particles*, is propagated in time, where the particles

- *explore* the state space under the effect of a *mutation* mechanism which mimics the evolution of the underlying process,
- and are *replicated* or *terminated*, under the effect of a *selection* mechanism which automatically concentrates the particles, i.e. the available computing power, into regions of interest of the state space.

In full generality, the underlying process is a discrete-time Markov chain, whose state space can be

finite, continuous, hybrid (continuous / discrete), graphical, constrained, time varying, pathwise, etc.,

the only condition being that it can easily be *simulated*.

In the special case of particle filtering, originally developed within the tracking community, the algorithms yield a numerical approximation of the optimal Bayesian filter, i.e. of the conditional probability distribution of the hidden state given the past observations, as a (possibly weighted) empirical probability distribution of the system of particles. In its simplest version, introduced in several different scientific communities under the name of *bootstrap filter* [38], *Monte Carlo filter* [43] or *condensation* (conditional density propagation) algorithm [42], and which historically has been the first algorithm to include a resampling step, the selection mechanism is governed by the likelihood function: at each time step, a particle is more likely to survive and to replicate at the next generation if it is consistent with the current observation. The algorithms also provide as a by–product a numerical approximation of the likelihood function, and of many other contrast functions for parameter estimation in hidden Markov models, such as the prediction error or the conditional least–squares criterion.

Particle methods are currently being used in many scientific and engineering areas

positioning, navigation, and tracking [39], [32], visual tracking [42], mobile robotics [33], [55], ubiquitous computing and ambient intelligence, sensor networks, risk evaluation and simulation of rare events [37], genetics, molecular simulation [34], etc.

Other examples of the many applications of particle filtering can be found in the contributed volume [22] and in the special issue of *IEEE Transactions on Signal Processing* devoted to *Monte Carlo Methods for Statistical Signal Processing* in February 2002, where the tutorial paper [23] can be found, and in the textbook [51] devoted to applications in target tracking. Applications of sequential Monte Carlo methods to other areas, beyond signal and image processing, e.g. to genetics, can be found in [48]. A recent overview can also be found in [25].

Particle methods are very easy to implement, since it is sufficient in principle to simulate independent trajectories of the underlying process. The whole problematic is multidisciplinary, not only because of the already mentioned diversity of the scientific and engineering areas in which particle methods are used, but also because of the diversity of the scientific communities which have contributed to establish the foundations of the field

target tracking, interacting particle systems, empirical processes, genetic algorithms (GA), hidden Markov models and nonlinear filtering, Bayesian statistics, Markov chain Monte Carlo (MCMC) methods.

These algorithms can be interpreted as numerical approximation schemes for Feynman–Kac distributions, a pathwise generalization of Gibbs–Boltzmann distributions, in terms of the weighted empirical probability distribution associated with a system of particles. This abstract point of view [30], [29], has proved to be extremely fruitful in providing a very general framework to the design and analysis of numerical approximation schemes, based on systems of branching and / or interacting particles, for nonlinear dynamical systems with values in the space of probability distributions, associated with Feynman–Kac distributions. Many asymptotic results have been proved as the number N of particles (sample size) goes to infinity, using techniques coming from applied probability (interacting particle systems, empirical processes [58]), see e.g. the survey article [30] or the textbooks [29], [28], and references therein

convergence in  $\mathbb{L}^p$ , convergence as empirical processes indexed by classes of functions, uniform convergence in time, see also [46], [47], central limit theorem, see also [44], [31], propagation of chaos, large deviations principle, etc.

The objective here is to systematically study the impact of the many algorithmic variants on the convergence results.

# 3.2. Multilevel splitting for rare event simulation

#### See 4.2, 5.1, and 5.2.

The estimation of the small probability of a rare but critical event, is a crucial issue in industrial areas such as

nuclear power plants, food industry, telecommunication networks, finance and insurance industry, air traffic management, etc.

In such complex systems, analytical methods cannot be used, and naive Monte Carlo methods are clearly unefficient to estimate accurately very small probabilities. Besides importance sampling, an alternate widespread technique consists in multilevel splitting [45], where trajectories going towards the critical set are given offsprings, thus increasing the number of trajectories that eventually reach the critical set. As shown in [6], the Feynman–Kac formalism of 3.1 is well suited for the design and analysis of splitting algorithms for rare event simulation.

**Propagation of uncertainty** Multilevel splitting can be used in static situations. Here, the objective is to learn the probability distribution of an output random variable Y = F(X), where the function F is only defined pointwise for instance by a computer programme, and where the probability distribution of the input random variable X is known and easy to simulate from. More specifically, the objective could be to compute the probability of the output random variable exceeding a threshold, or more generally to evaluate the cumulative distribution function of the output random variable for different output values. This problem is characterized by the lack of an analytical expression for the function, the computational cost of a single pointwise evaluation of the function, which means that the number of calls to the function should be limited as much as possible,

and finally the complexity and / or unavailability of the source code of the computer programme, which makes any modification very difficult or even impossible, for instance to change the model as in importance sampling methods.

The key issue is to learn as fast as possible regions of the input space which contribute most to the computation of the target quantity. The proposed splitting methods consists in (i) introducing a sequence of intermediate regions in the input space, implicitly defined by exceeding an increasing sequence of thresholds or levels, (ii) counting the fraction of samples that reach a level given that the previous level has been reached already, and (iii) improving the diversity of the selected samples, usually with an artificial Markovian dynamics for the input variable. In this way, the algorithm learns

- the transition probability between successive levels, hence the probability of reaching each intermediate level,
- and the probability distribution of the input random variable, conditionned on the output variable reaching each intermediate level.

A further remark, is that this conditional probability distribution is precisely the optimal (zero variance) importance distribution needed to compute the probability of reaching the considered intermediate level.

**Rare event simulation** To be specific, consider a complex dynamical system modelled as a Markov process, whose state can possibly contain continuous components and finite components (mode, regime, etc.), and the objective is to compute the probability, hopefully very small, that a critical region of the state space is reached by the Markov process before a final time T, which can be deterministic and fixed, or random (for instance the time of return to a recurrent set, corresponding to a nominal behaviour).

The proposed splitting method consists in (i) introducing a decreasing sequence of intermediate, more and more critical, regions in the state space, (ii) counting the fraction of trajectories that reach an intermediate region before time T, given that the previous intermediate region has been reached before time T, and (iii) regenerating the population at each stage, through resampling. In addition to the non-intrusive behaviour of the method, the splitting methods make it possible to learn the probability distribution of typical critical trajectories, which reach the critical region before final time T, an important feature that methods based on importance sampling usually miss. Many variants have been proposed, whether

- the branching rate (number of offsprings allocated to a successful trajectory) is fixed, which allows for depth-first exploration of the branching tree, but raises the issue of controlling the population size,
- the population size is fixed, which requires a breadth-first exploration of the branching tree, with random (multinomial) or deterministic allocation of offsprings, etc.

Just as in the static case, the algorithm learns

- the transition probability between successive levels, hence the probability of reaching each intermediate level,
- and the entrance probability distribution of the Markov process in each intermediate region.

Contributions have been given to

- minimizing the asymptotic variance, obtained through a central limit theorem, with respect to the shape of the intermediate regions (selection of the importance function), to the thresholds (levels), to the population size, etc.
- controlling the probability of extinction (when not even one trajectory reaches the next intermediate level),
- designing and studying variants suited for hybrid state space (resampling per mode, marginalization, mode aggregation),

and in the static case, to

• minimizing the asymptotic variance, obtained through a central limit theorem, with respect to intermediate levels, to the Metropolis kernel introduced in the mutation step, etc.

A related issue is global optimization. Indeed, the difficult problem of finding the set M of global minima of a real-valued function V can be replaced by the apparently simpler problem of sampling a population from a probability distribution depending on a small parameter, and asymptotically supported by the set M as the small parameter goes to zero. The usual approach here is to use the cross-entropy method [52], [27], which relies on learning the optimal importance distribution within a prescribed parametric family. On the other hand, multilevel splitting methods could provide an alternate nonparametric approach to this problem.

# 3.3. Statistical learning: pattern recognition and nonparametric regression

In pattern recognition and statistical learning, also known as machine learning, nearest neighbor (NN) algorithms are amongst the simplest but also very powerful algorithms available. Basically, given a training set of data, i.e. an N-sample of i.i.d. object-feature pairs, with real-valued features, the question is how to generalize, that is how to guess the feature associated with any new object. To achieve this, one chooses some integer k smaller than N, and takes the mean-value of the k features associated with the k objects that are nearest to the new object, for some given metric.

In general, there is no way to guess exactly the value of the feature associated with the new object, and the minimal error that can be done is that of the Bayes estimator, which cannot be computed by lack of knowledge of the distribution of the object–feature pair, but the Bayes estimator can be useful to characterize the strength of the method. So the best that can be expected is that the NN estimator converges, say when the sample size N grows, to the Bayes estimator. This is what has been proved in great generality by Stone [53] for the mean square convergence, provided that the object is a finite–dimensional random variable, the feature is a square–integrable random variable, and the ratio k/N goes to 0. Nearest neighbor estimator is not the only local averaging estimator with this property, but it is arguably the simplest.

The asymptotic behavior when the sample size grows is well understood in finite dimension, but the situation is radically different in general infinite dimensional spaces, when the objects to be classified are functions, images, etc.

Nearest neighbor classification in infinite dimension In finite dimension, the k-nearest neighbor classifier is universally consistent, i.e. its probability of error converges to the Bayes risk as N goes to infinity, whatever the joint probability distribution of the pair, provided that the ratio k/N goes to zero. Unfortunately, this result is no longer valid in general metric spaces, and the objective is to find out reasonable sufficient conditions for the weak consistency to hold. Even in finite dimension, there are exotic distances such that the nearest neighbor does not even get closer (in the sense of the distance) to the point of interest, and the state space needs to be complete for the metric, which is the first condition. Some regularity on the regression function is required next. Clearly, continuity is too strong because it is not required in finite dimension, and a weaker form of regularity is assumed. The following consistency result has been obtained: if the metric space is separable and if some Besicovich condition holds, then the nearest neighbor classifier is weakly consistent. Note that the Besicovich condition is always fulfilled in finite dimensional vector spaces (this result is called the Besicovich theorem), and that a counterexample [4] can be given in an infinite dimensional space with a Gaussian measure (in this case, the nearest neighbor classifier is clearly nonconsistent). Finally, a simple example has been found which verifies the Besicovich condition with a noncontinuous regression function.

**Rates of convergence of the functional** k-nearest neighbor estimator Motivated by a broad range of potential applications, such as regression on curves, rates of convergence of the k-nearest neighbor estimator of the regression function, based on N independent copies of the object-feature pair, have been investigated when the object is in a suitable ball in some functional space. Using compact embedding theory, explicit and general finite sample bounds can be obtained for the expected squared difference between the k-nearest neighbor estimator and the Bayes regression function, in a very general setting. The results have also been particularized to classical function spaces such as Sobolev spaces, Besov spaces and reproducing kernel Hilbert spaces. The rates obtained are genuine nonparametric convergence rates, and up to our knowledge the first of their kind for k-nearest neighbor regression.

This topic has produced several theoretical advances [1], [2] in collaboration with Gérard Biau (université Pierre et Marie Curie). A few possible target application domains have been identified in

- the statistical analysis of recommendation systems,
- the design of reduced-order models and analog samplers,

that would be a source of interesting problems.

# **4.** Application Domains

# 4.1. Localisation, navigation and tracking

#### See 6.1.

Among the many application domains of particle methods, or interacting Monte Carlo methods, ASPI has decided to focus on applications in localisation (or positioning), navigation and tracking [39], [32], which already covers a very broad spectrum of application domains. The objective here is to estimate the position (and also velocity, attitude, etc.) of a mobile object, from the combination of different sources of information, including

- a prior dynamical model of typical evolutions of the mobile, such as inertial estimates and prior model for inertial errors,
- measurements provided by sensors,
- and possibly a digital map providing some useful feature (terrain altitude, power attenuation, etc.) at each possible position.

In some applications, another useful source of information is provided by

• a map of constrained admissible displacements, for instance in the form of an indoor building map,

which particle methods can easily handle (map-matching). This Bayesian dynamical estimation problem is also called filtering, and its numerical implementation using particle methods, known as particle filtering, has been introduced by the target tracking community [38], [51], which has already contributed to many of the most interesting algorithmic improvements and is still very active, and has found applications in

target tracking, integrated navigation, points and / or objects tracking in video sequences, mobile robotics, wireless communications, ubiquitous computing and ambient intelligence, sensor networks, etc.

ASPI is contributing (or has contributed recently) to several applications of particle filtering in positioning, navigation and tracking, such as geolocalisation and tracking in a wireless network, terrain–aided navigation, and data fusion for indoor localisation.

# 4.2. Rare event simulation

#### See 3.2, 5.1, and 5.2.

Another application domain of particle methods, or interacting Monte Carlo methods, that ASPI has decided to focus on is the estimation of the small probability of a rare but critical event, in complex dynamical systems. This is a crucial issue in industrial areas such as

nuclear power plants, food industry, telecommunication networks, finance and insurance industry, air traffic management, etc.

In such complex systems, analytical methods cannot be used, and naive Monte Carlo methods are clearly unefficient to estimate accurately very small probabilities. Besides importance sampling, an alternate widespread technique consists in multilevel splitting [45], where trajectories going towards the critical set are given offsprings, thus increasing the number of trajectories that eventually reach the critical set. This approach not only makes it possible to estimate the probability of the rare event, but also provides realizations of the random trajectory, given that it reaches the critical set, i.e. provides realizations of typical critical trajectories, an important feature that methods based on importance sampling usually miss. ASPI is contributing (or has contributed recently) to several applications of multilevel splitting for rare event simulation, such as risk assessment in air traffic management, detection in sensor networks, and protection of digital documents.

# 5. New Results

# 5.1. Central limit theorem for adaptive multilevel splitting

Participants: Frédéric Cérou, Arnaud Guyader, Mathias Rousset.

See 3.2, and 4.2.

This is a collaboration with Bernard Delyon (université de Rennes 1).

Fleming–Viot type particle systems represent a classical way to approximate the distribution of a Markov process with killing, given that it is still alive at a final deterministic time. In this context, each particle evolves independently according to the law of the underlying Markov process until its killing, and then branches instantaneously on another randomly chosen particle. While the consistency of this algorithm in the large population limit has been recently studied in several articles, our purpose here is to prove central limit theorems under very general assumptions. For this, we only suppose that the particle system does not explode in finite time, and that the jump and killing times have atomless distributions. In particular, this includes the case of elliptic diffusions with hard killing.

# 5.2. Adaptive multilevel splitting for Monte Carlo particle transport

Participant: Mathias Rousset.

See 3.2, and 4.2.

Simulation of neutron transport with Monte Carlo methods is a central issue in order to assess the aging of french nucelar plants.

In [49], we propose an alternative version of the AMS (adaptive multilevel splitting) algorithm, adapted for the first time to the field of particle tranport. Within this context, it can be used to build an unbiased estimator of any quantity associated with particle tracks, such as flux, reaction rates or even non–Boltzmann tallies. Furthermore, the effciency of the AMS algorithm is shown not to be very sensitive to variations of its input parameters, which makes it capable of significant variance reduction without requiring extended user effort.

## 5.3. Weak overdamped limit theorem for Langevin processes

Participant: Mathias Rousset.

This is a collaboration with Pierre-André Zitt (université Paris Est Marne-la-Vallée).

The Langevin stochastic process is the main model used in molecular dynamics simulation, for instance for the simulation of reactive trajectories of bio-chemical systems with rare event techniques.

In [21], we prove convergence in distribution of Langevin processes in the overdamped diffusion asymptotics. The proof relies on the classical perturbed test function (or corrector) method, which is used both to show tightness in path space, and to identify the extracted limit with a martingale problem. The result holds assuming the continuity of the gradient of the potential energy, and a mild control of the initial kinetic energy.

## 5.4. Particle–Kalman filter for structural health monitoring

Participant: Frédéric Cérou.

This is a joint work with EPI I4S (Inria Rennes-Bretagne Atlantique).

Standard filtering techniques for structural parameter estimation assume that the input force either is known exactly or can be replicated using a known white Gaussian model. Unfortunately for structures subjected to seismic excitation, the input time history is unknown and also no previously known representative model is available. This invalidates the aforementioned idealization. To identify seismic induced damage in such structures using filtering techniques, a novel algorithm is proposed to estimate the force as additional state in parallel to the system parameters. Two concurrent filters are employed for parameters and force respectively. For the parameters, interacting particle–Kalman filter is employed targeting systems with correlated noise. Alongside a second filter is employed to estimate the seismic force acting on the structure. The proposal is numerically validated on a sixteen degrees–of–freedom mass–spring–damper system. The estimation results confirm the applicability of the proposed algorithm.

In another work, the same approach has been used for varying system parameters with correlated state and observation noise. The idea is to nest a bank of linear KFs (Kalman filters) for state estimation within a PF (particle filter) environment that estimates the parameters. This facilitates employing relatively less expensive linear KF for linear state estimation problem while costly PF is employed only for parameter estimation. Additionally, the proposed algorithm also takes care of those systems for which system and measurement noises are not uncorrelated as it is commonly idealized in standard filtering algorithms. As an example, for mechanical systems under ambient vibration it happens when acceleration response is considered as measurement. Thus the process and measurement noise in these system descriptions are obviously correlated. For this, an improved description for the Kalman gain is developed. Further, to enhance the consistency of particle filtering based parameter estimation involving high dimensional parameter space, a new temporal evolution strategy for the particles is defined. This strategy aims at restricting the solution from diverging (up to the point of no return) because of an isolated event of infeasible estimation which is very much likely especially when dealing with high dimensional parameter space.

#### 5.5. Reduced modeling of unknown trajectories

Participant: Patrick Héas.

This is a collaboration with Cédric Herzet (EPI FLUMINANCE, Inria Rennes-Bretagne Atlantique)

In [12], we deal with model order reduction of parametrical dynamical systems. We consider the specific setup where the distribution of the system's trajectories is unknown but the following two sources of information are available: (*i*) some "rough" prior knowledge on the system's realisations, and (*ii*) a set of "incomplete" observations of the system's trajectories. We propose a Bayesian methodological framework to build reduced–order models (ROMs) by exploiting these two sources of information.

We emphasise that complementing the prior knowledge with the collected data provably enhances the knowledge of the distribution of the system's trajectories. We then propose an implementation of the proposed methodology based on Monte Carlo methods. In this context, we show that standard ROM learning techniques, such as proper orthogonal decomposition (POD) or dynamic mode decomposition (DMD), can be revisited and recast within the probabilistic framework considered in this work. We illustrate the performance of the proposed approach by numerical results obtained for a standard geophysical model.

# 5.6. Model reduction from partial observations

Participant: Patrick Héas.

This is a collaboration with Angélique Drémeau (ENSTA Bretagne, Brest) and Cédric Herzet (EPI FLUMI-NANCE, Inria Rennes–Bretagne Atlantique)

In [11], we deal with model-order reduction of parametric partial differential equations (PPDE). More specifically, we consider the problem of finding a good approximation subspace of the solution manifold of the PPDE when only partial information on the latter is available. We assume that two sources of information are available: i) a "rough" prior knowledge, taking the form of a manifold containing the target solution manifold, and ii) partial linear measurements of the solutions of the PPDE (the term partial refers

to the fact that observation operator cannot be inverted). We provide and study several tools to derive good approximation subspaces from these two sources of information. We first identify the best worst-case performance achievable in this setup and propose simple procedures to approximate the corresponding optimal approximation subspace. We then provide, in a simplified setup, a theoretical analysis relating the achievable reduction performance to the choice of the observation operator and the prior knowledge available on the solution manifold.

# 5.7. Low–rank dynamic mode decomposition: optimal solution in polynomial time

Participant: Patrick Héas.

This is a collaboration with Cédric Herzet (EPI FLUMINANCE, Inria Rennes-Bretagne Atlantique)

The works [15] and [41] study the linear approximation of high–dimensional dynamical systems using low-rank dynamic mode decomposition (DMD). Searching this approximation in a data–driven approach can be formalised as attempting to solve a low-rank constrained optimisation problem. This problem is non–convex and state–of–the–art algorithms are all sub–optimal. We show that there exists a closed-form solution, which can be computed in polynomial time, and characterises the  $\ell_2$ –norm of the optimal approximation error. The theoretical results serve to design low–complexity algorithms building reduced models from the optimal solution, based on singular value decomposition or low–rank DMD. The algorithms are evaluated by numerical simulations using synthetic and physical data benchmarks.

# 5.8. Optimal kernel-based dynamic mode decomposition

Participant: Patrick Héas.

This is a collaboration with Cédric Herzet (EPI FLUMINANCE, Inria Rennes-Bretagne Atlantique)

The state–of–the–art algorithm known as kernel-based dynamic mode decomposition (K–DMD) provides a sub–optimal solution to the problem of reduced modeling of a dynamical system based on a finite approximation of the Koopman operator. It relies on crude approximations and on restrictive assumptions. The purpose of the work in [20] is to propose a kernel–based algorithm solving exactly this low–rank approximation problem in a general setting.

## 5.9. Non parametric state–space model for missing–data imputation

Participants: Thi Tuyet Trang Chau, François Le Gland, Valérie Monbet, Mathias Rousset.

This is a collaboration with Pierre Ailliot (université de Bretagne Occidentale, Brest), Ronan Fablet and Pierre Tandéo (Télécom Bretagne, Brest), Anne Cuzol (université de Bretagne Sud, Vannes) and Bernard Chapron (IFREMER, Brest).

Missing data are present in many environmental data–sets and this work aims at developing a general method for imputing them. State–space models (SSM) have already extensively been used in this framework. The basic idea consists in introducing the true environmental process, which we aim at reconstructing, as a latent process and model the data available at neighboring sites in space and/or time conditionally to this latent process. A key input of SSMs is a stochastic model which describes the temporal evolution of the environmental process of interest. In many applications, the dynamic is complex and can hardly be described using a tractable parametric model. Here we investigate a data-driven method where the dynamical model is learned using a non-parametric approach and historical observations of the environmental process of interest. From a statistical point of view, we will address various aspects related to SSMs in a non–parametric framework. First we will discuss the estimation of the filtering and smoothing distributions, that is the distribution of the latent space given the observations, using sequential Monte Carlo approaches in conjunction with local linear regression. Then, a more difficult and original question consists in building a non–parametric estimate of the dynamics which takes into account the measurement errors which are present in historical data. We will propose an EM–like algorithm where the historical data are corrected recursively. The methodology will be illustrated and validated on an univariate toy example.

# 6. Bilateral Contracts and Grants with Industry

# 6.1. Bilateral grants with industry

See <mark>4.1</mark>.

#### 6.1.1. Hybrid indoor navigation — PhD project at CEA LETI

Participants: François Le Gland, Kersane Zoubert-Ousseni.

This is a collaboration with Christophe Villien (CEA LETI, Grenoble).

The issue here is user localization, and more generally localization-based services (LBS). This problem is addressed by GPS for outdoor applications, but no such general solution has been provided so far for indoor applications. The desired solution should rely on sensors that are already available on smartphones and other tablet computers. Inertial solutions that use MEMS (microelectromechanical system, such as accelerometer, magnetometer, gyroscope and barometer) are already studied at CEA. An increase in performance should be possible, provided these data are combined with other available data: map of the building, WiFi signal, modeling of perturbations of the magnetic field, etc. To be successful, advanced data fusion techniques should be used, such as particle filtering and the like, to take into account displacement constraints due to walls in the building, to manage several possible trajectories, and to deal with rather heterogeneous information (map, radio signals, sensor signals).

The main objective of this thesis is to design and tune localization algorithms that will be tested on platforms already available at CEA. Special attention is paid to particle smoothing and particle MCMC algorithms, to exploit some very precise information available at special time instants, e.g. when the user is clearly localized near a landmark point.

In some applications, real time estimation of the trajectory is not needed, and a post processing framework may provide a better estimation of this trajectory. In [57], we present and compare three different algorithms to improve a real time trajectory estimation. Actually, two different smoothing algorithms and the Viterbi algorithm are implemented and evaluated. These methods improve the regularity of the estimated trajectory by reducing switches between hypotheses.

Post processing indoor navigation is interesting, for example to develop crowdsourcing analysis. The post processing framework allows to provide a better estimation than in a real time framework. The main contribution of [17] is to present a piecewise parametrization using IMU (inertial measurement unit) and RSS (received signal strength) measurements only, which lead to an optimization problem. A Levenberg–Marquardt algorithm improved with simulated annealing and an adjustment of RSS measurements data leads to a good estimation (55% of the error less than 5 meters) of the trajectory.

## 6.1.2. Bayesian tracking from raw data — CIFRE grant with DCNS Nantes

Participants: François Le Gland, Audrey Cuillery.

This is a collaboration with Dann Laneuville (DCNS Nantes).

After the introduction of MHT (multi–hypothesis tracking) techniques in the nineties, multitarget tracking has recently seen promising developpments with the introduction of new algorithms such as the PHD (probability hypothesis density) filter [50], [56] or the HISP (hypothesised filter for independent stochastic populations) filter [40]. These techniques provide a unified multitarget model in a Bayesian framework [54], which makes it possible to design recursive estimators of a *multitarget probability density*. Two main approaches can be used here: sequential Monte Carlo (SMC, also kown as particle filtering), and Gaussian mixture (GM). A third approach, based on discretizing the state–space in a possibly adaptive way, could also be considered despite its larger computational load. These methods are well studied and provide quite good results for *contact output* data, which correspond to regularly spaced measurements of targets with a large SNR (signal–to–noise ratio). Here, the data is processed (compared with a detection threshold) in each resolution cell of the sensor, so as to provide a list of detections at a given time instant. Among these methods, the HISP filter has the best performance/computational cost ratio.

However, these classical methods are unefficient for targets with a low SNR, e.g. targets in far range or small targets with a small detection probability. For such targets, preprocessing (thresholding) the data is not a good idea, and a much better idea is to feed a tracking algorithm with the raw *sensor output* data directly. These new methods [24] require a precise modeling of the sensor physics and a direct access to the radar (or the sonar) raw data, i.e. to the signal intensity level in each azimuth/range cell. Note that these new methods seem well suited to new types of sensors such as lidar, since manufacturers do not integrate a detection module and do provide raw images of the signal intensity level in each azimuth/range cell.

The objective of the thesis is to study and design a tracking algorithm using raw data, and to implement it on radar (or sonar, or lidar) real data.

# 7. Partnerships and Cooperations

# 7.1. Regional initiatives

# 7.1.1. Stochastic Model-Data Coupled Representations for the Upper Ocean Dynamics (SEACS) — inter labex project

Participants: François Le Gland, Valérie Monbet.

January 2015 to December 2017.

This is a joint research initiative supported by the three labex active in Brittany, CominLabs (Communication and Information Sciences Laboratory), Lebesgue (Centre de Mathématiques Henri Lebesgue) and LabexMER (Frontiers in Marine Research).

This project aims at exploring novel statistical and stochastic methods to address the emulation, reconstruction and forecast of fine–scale upper ocean dynamics. The key objective is to investigate new tools and methods for the calibration and implementation of novel sound and efficient oceanic dynamical models, combining

- recent advances in the theoretical understanding, modeling and simulation of upper ocean dynamics,
- and mass of data routinely available to observe the ocean evolution.

In this respect, the emphasis will be given to stochastic frameworks to encompass multi-scale/multi-source approaches and benefit from the available observation and simulation massive data. The addressed scientific questions constitute basic research issues at the frontiers of several disciplines. It crosses in particular advanced data analysis approaches, physical oceanography and stochastic representations. To develop such an interdisciplinary initiative, the project gathers a set of research groups associated with these different scientific domains, which have already proven for several years their capacities to interact and collaborate on topics related to oceanic data and models. This project will place Brittany with an innovative and leading expertise at the frontiers of computer science, statistics and oceanography. This transdisciplinary research initiative is expected to resort to significant advances challenging the current thinking in computational oceanography.

# 7.2. National initiatives

# 7.2.1. Computational Statistics and Molecular Simulation (COSMOS) — ANR challenge Information and Communication Society

Participant: Frédéric Cérou.

#### Inria contract ALLOC 9452 — January 2015 to December 2017.

The COSMOS project aims at developing numerical techniques dedicated to the sampling of high–dimensional probability measures describing a system of interest. There are two application fields of interest: computational statistical physics (a field also known as molecular simulation), and computational statistics. These two fields share some common history, but it seems that, in view of the quite recent specialization of the scientists and the techniques used in these respective fields, the communication between molecular simulation and computational statistics is not as intense as it should be.

We believe that there are therefore many opportunities in considering both fields at the same time: in particular, the adaption of a successful simulation technique from one field to the other requires first some abstraction process where the features specific to the original field of application are discarded and only the heart of the method is kept. Such a cross–fertilization is however only possible if the techniques developed in a specific field are sufficiently mature: this is why some fundamental studies specific to one of the application fields are still required. Our belief is that the embedding in a more general framework of specific developments in a given field will accelerate and facilitate the diffusion to the other field.

## 7.2.2. Advanced Geophysical Reduced–Order Model Construction from Image Observations (GERONIMO) — ANR programme Jeunes Chercheuses et Jeunes Chercheurs Participant: Patrick Héas.

#### Inria contract ALLOC 8102 — March 2014 to February 2018.

The GERONIMO project aims at devising new efficient and effective techniques for the design of geophysical reduced–order models (ROMs) from image data. The project both arises from the crucial need of accurate low–order descriptions of highly–complex geophysical phenomena and the recent numerical revolution which has supplied the geophysical scientists with an unprecedented volume of image data. Our research activities are concerned by the exploitation of the huge amount of information contained in image data in order to reduce the uncertainty on the unknown parameters of the models and improve the reduced–model accuracy. In other words, the objective of our researches to process the large amount of incomplete and noisy image data daily captured by satellites sensors to devise new advanced model reduction techniques. The construction of ROMs is placed into a probabilistic Bayesian inference context, allowing for the handling of uncertainties associated to image measurements and the characterization of parameters of the reduced dynamical system.

## 7.3. European initiatives

# 7.3.1. Molecular Simulation: Modeling, Algorithms and Mathematical Analysis (MSMaths) — ERC Consolidator Grant

Participant: Mathias Rousset.

#### January 2014 to December 2019.

PI: Tony Lelièvre, Civil Engineer in Chief, Ecole des Ponts Paris-Tech.

Note that 1/3 of Mathias Rousset research activities are held within the MSMath ERC project.

With the development of large-scale computing facilities, simulations of materials at the molecular scale are now performed on a daily basis. The aim of these simulations is to understand the macroscopic properties of matter from a microscopic description, for example, its atomistic configuration.

In order to make these simulations efficient and precise, mathematics have a crucial role to play. Indeed, specific algorithms have to be used in order to bridge the time and space scales between the atomistic level and the macroscopic level. The objective of the MSMath ERC project is thus to develop and study efficient algorithms to simulate high–dimensional systems over very long times. These developments are done in collaboration with physicists, chemists and biologists who are using these numerical methods in an academic or industrial context.

In particular, we are developping mathematical tools at the interface between the analysis of partial differential equations and stochastic analysis in order to characterize and to quantify the metastability of stochastic processes. Metastability is a fundamental concept to understand the timescale separation between the microscopic model and the macroscopic world. Many algorithms which aim at bridging the timescales are built using this timescale separation.

# 7.3.2. Design of Desalination Systems Based on Optimal Usage of Multiple Renewable Energy Sources (DESIRES) — ERANETMED NEXUS-14-049

Participant: Valérie Monbet.

#### January 2016 to December 2018.

This project is funded by the ERA–NET Initiative ERANETMED (Euro–Mediterranean Cooperation through ERA–NET Joint Activities and Beyond). It is a collaboration with Greece, Tunisia and Marocco, coordinated by Technical University of Crete (TUC). The French staff includes: Pierre Ailliot (Université de Bretagne Occidentale, Brest), Denis Allard (INRA Avignon), Anne Cuzol (Université de Bretagne Sud, Vannes), Christophe Maisondieu (IFREMER Brest) and Valérie Monbet.

The aim of **DESIRES** is to develop an Internet–based, multi–parametric electronic platform for optimum design of desalination plants, supplied by renewable energy sources (RES). The platform will rely upon (i) a solar, wind and wave energy potential database, (ii) existing statistical algorithms for processing energy-related data, (iii) information regarding the inter-annual water needs, (iv) a database with the technical characteristics of desalination plant units and the RES components, and (v) existing algorithms for cost effective design, optimal sizing and location selection of desalination plants.

## 7.4. International initiatives

#### 7.4.1. Rare event simulation in epidemiology — PhD project at université de Ziguinchor

Participants: Ramatoulaye Dabo, François Le Gland.

This is the subjet of the PhD project of Ramatoulaye Dabo (université Assane Seck de Ziguinchor and université de Rennes 1).

The question here is to develop adaptive multilevel splitting algorithms for models that are commonly used in epidemiology, such as SIR (susceptible, infectious, recovered) models [26], or more complex compartmental models. A significant advantage of adaptive multilevel splitting is its robustness, since it does not require too much knowledge about the behavior of the system under study. An interesting challenge would be to understand how to couple the algorithm with numerically efficient simulation methods such as  $\tau$ -leaping [35]. Complexity bounds and estimation error bounds could also be studied.

# 7.5. International research visitors

#### 7.5.1. Visits to international teams

Patrick Héas has been invited to present his work on 3D wind field reconstruction by infrared sounding, at EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites) in Darmstadt in February 2017.

# 8. Dissemination

## 8.1. Promoting scientific activities

#### 8.1.1. Scientific events organisation

Valérie Monbet has co-organized the workshop and summer school on Data Science and Environment, held in Brest in July 2017. The conference gathered researchers that have an expertise in one of the two areas (data science, environmental data) and some interest for the other. Its main goal was to explore the fruitful interplay between the two areas, and ultimately to help create new connections and collaborations between the scientific communities involved. Another objective was to propose some high level courses and practices at the interaction of these two areas.

#### 8.1.2. Participation in workshops, seminars, lectures, etc.

In addition to presentations with a publication in the proceedings, which are listed at the end of the document, members of ASPI have also given the following presentations.

Frédéric Cérou has given an invited talk on the convergence of adaptive multilevel splitting at the workshop Quasistationary Distributions: Analysis and Simulation held in Paderborn in September 2017.

Patrick Héas has presented his joint work with Mamadou Lamarana Diallo and Cédric Herzet (EPI FLUMI-NANCE, Inria Rennes–Bretagne Atlantique) on model reduction with "multi-space" prior information, at the European Conference on Numerical Mathematics and Advanced Applications (ENUMATH), held in Voss, Norway, in September 2017.

Thi Tuyet Trang Chau has presented her work on non parametric state–space model for missing–data imputation, at the workshop on Data Science and Environment, held in Brest in July 2017.

#### 8.1.3. Research administration

François Le Gland is a member of the *conseil d'UFR* of the department of mathematics of université de Rennes 1. He is also a member of the *conseil scientifique* for the EDF/Inria scientific partnership.

Valérie Monbet is a member of both the *comité de direction* and the *conseil* of IRMAR (institut de recherche mathématiques de Rennes, UMR 6625). She is also the deputy head of the department of mathematics of université de Rennes 1, where she is a member of both the *conseil scientifique* and the *conseil d'UFR*.

# 8.2. Teaching, supervision, thesis committees

## 8.2.1. Teaching

Patrick Héas gives a course on Monte Carlo simulation methods in image analysis, at université de Rennes 1, within the SISEA (signal, image, systèmes embarqués, automatique) track of the master in electronical engineering and telecommunications.

François Le Gland gives

- a 2nd year course on introduction to stochastic differential equations, at INSA (institut national des sciences appliquées) Rennes, within the GM/AROM (risk analysis, optimization and modeling) major in mathematical engineering,
- a 3rd year course on Bayesian filtering and particle approximation, at ENSTA (école nationale supérieure de techniques avancées), Palaiseau, within the statistics and control module,
- a 3rd year course on linear and nonlinear filtering, at ENSAI (école nationale de la statistique et de l'analyse de l'information), Ker Lann, within the statistical engineering track,
- a course on Kalman filtering and hidden Markov models, at université de Rennes 1, within the SISEA (signal, image, systèmes embarqués, automatique, école doctorale MATISSE) track of the master in electronical engineering and telecommunications,
- and a 3rd year course on hidden Markov models, at Télécom Bretagne, Brest.

Valérie Monbet gives

- a course on machine learning for biology at université de Rennes 1, within
  - the G2B (genetics, genomics, biochemistry) track of the master in molecular and cellular biology,
  - the MODE (modélisation en écologie) track of the master in biodiversity, ecology, evolution
  - and the master in scientific computing and modelling,
- a course on machine learning for environmental data, at the summer school on Data Science and Environment, held in Brest in July 2017,
- a course on graphical models at université de Rennes 1, within the master on applied mathematics and statistics,
- a course on MATLAB at université de Rennes 1, within the master in economics and financial engineering.

#### 8.2.2. Supervision

François Le Gland and Valérie Monbet are jointly supervising one PhD student

• Thi Tuyet Trang Chau, provisional title: *Non parametric filtering for Metocean multi–source data fusion*, université de Rennes 1, started in October 2015, expected defense in October 2018, funding: Labex Lebesgue grant and Brittany council grant, co–direction: Pierre Ailliot (université de Bretagne Occidentale, Brest).

François Le Gland is supervising three other PhD students

- Kersane Zoubert–Ousseni, provisional title: *Particle filters for hybrid indoor navigation with smart-phones*, université de Rennes 1, started in December 2014, expected defense in 2017, funding: CEA grant, co–direction: Christophe Villien (CEA LETI, Grenoble),
- Audrey Cuillery, provisional title: *Bayesian tracking from raw data*, université du Sud Toulon Var, started in April 2016, expected defense in 2019, funding: CIFRE grant with DCNS, co-direction: Claude Jauffret (université du Sud Toulon Var) and Dann Laneuville (DCNS, Nantes).
- Ramatoulaye Dabo, provisional title: *Rare event simulation in epidemiology*, université Assane Seck de Ziguinchor (Senegal) and université de Rennes 1, started in September 2015, expected defense in 2018, co-direction: Alassane Diedhiou (université Assane Seck de Ziguinchor).

Valérie Monbet is supervising two other PhD students

- Audrey Poterie, provisional title: *Régression d'une variable ordinale par des données longitudinales de grande dimension : application à la modélisation des effets secondaires suite à un traitement par radiothérapie*, université de Rennes 1, started in October 2015, expected defense in 2018, funding: INSA grant, co-direction: Jean-François Dupuy (INSA Rennes) and Laurent Rouvière (université de Haute Bretagne, Rennes).
- Marie Morvan, provisional title: *Modèles de régression pour données fonctionnelles. Application à la modélisation de données de spectrométrie dans le proche infra rouge*, université de Rennes 1, started in October 2016, expected defense in 2019, funding: MESR grant, co-direction: Joyce Giacofci (université de Haute Bretagne, Rennes) and Olivier Sire (université de Bretagne Sud, Vannes).

Mathias Rousset is supervising one PhD student

• Yushun Xu, provisional title: *Variance reduction of overdamped Langevin dynamics simulation*, université Paris-Est, started in October 2015, expected defense in 2018, co-direction: Pierre-André Zitt (université Paris-Est).

Patrick Héas has been supervising two post-doctoral fellows

- Hassan Maatouk, title: *Compressing the model by exploiting observations*, EPI ASPI, Inria Rennes–Bretagne Atlantique, started in September 2016, ended in September 2017, funding: ANR GERONIMO, co–supervision: Cédric Herzet (EPI FLUMINANCE, Inria Rennes–Bretagne Atlantique).
- Mamadou Lamarana Diallo, title: *Model reduction with "multi-space" prior information*, EPI FLUMINANCE, Inria Rennes–Bretagne Atlantique, started in October 2016, ended in October 2017, funding: ANR GERONIMO, co–supervision: Cédric Herzet (EPI FLUMINANCE, Inria Rennes–Bretagne Atlantique).

## 8.2.3. Thesis committees

François Le Gland has been a member of the committee for the HDR of Christian Musso (université du Sud, Toulon).

Mathias Rousset has been a member of the committee for the PhD thesis of Gérôme Faure (CERMICS Ecole des Ponts Paris-Tech and CEA DAM, advisor: Gabriel Stoltz and Jean–Bernard Maillet).

# 9. Bibliography

# Major publications by the team in recent years

- G. BIAU, F. CÉROU, A. GUYADER. On the rate of convergence of the bagged nearest neighbor estimate, in "Journal of Machine Learning Research", February 2010, vol. 11, p. 687–712, http://jmlr.csail.mit.edu/papers/ v11/biau10a.html.
- [2] G. BIAU, F. CÉROU, A. GUYADER. On the rate of convergence of the functional k-nearest neighbor estimates, in "IEEE Transactions on Information Theory", April 2010, vol. IT-56, n<sup>o</sup> 4, p. 2034–2040, http://dx.doi.org/ 10.1109/TIT.2010.2040857.
- [3] G. BIAU, F. CÉROU, A. GUYADER. New insights into approximate Bayesian computation, in "Annales de l'Institut Henri Poincaré, Probabilités et Statistiques", February 2015, vol. 51, nº 1, p. 376–403, http://dx.doi. org/10.1214/13-AIHP590.
- [4] F. CÉROU, A. GUYADER. Nearest neighbor classification in infinite dimension, in "ESAIM : Probability and Statistics", 2006, vol. 10, p. 340–355, http://dx.doi.org/10.1051/ps:2006014.
- [5] F. CÉROU, A. GUYADER. Adaptive multilevel splitting for rare event analysis, in "Stochastic Analysis and Applications", March 2007, vol. 25, n<sup>o</sup> 2, p. 417–443, http://dx.doi.org/10.1080/07362990601139628.
- [6] F. CÉROU, P. DEL MORAL, F. LE GLAND, P. LEZAUD. Genetic genealogical models in rare event analysis, in "ALEA, Latin American Journal of Probability and Mathematical Statistics", 2006, vol. 1, p. 181–203, Paper 01–08.
- [7] A. GUYADER, N. HENGARTNER, É. MATZNER–LØBER.Simulation and estimation of extreme quantiles and extreme probabilities, in "Applied Mathematics & Optimization", October 2011, vol. 64, n<sup>o</sup> 2, p. 171–196, http://dx.doi.org/10.1007/s00245-011-9135-z.
- [8] F. LE GLAND, V. MONBET, V.-D. TRAN.Large sample asymptotics for the ensemble Kalman filter, in "The Oxford Handbook of Nonlinear Filtering", Oxford, D. O. CRISAN, B. L. ROZOVSKII (editors), Oxford University Press, 2011, chap. 22, p. 598–631.
- [9] F. LE GLAND, N. OUDJANE. A sequential algorithm that keeps the particle system alive, in "Stochastic Hybrid Systems : Theory and Safety Critical Applications", Berlin, H. A. P. BLOM, J. LYGEROS (editors), Lecture Notes in Control and Information Sciences, Springer–Verlag, 2006, n<sup>o</sup> 337, p. 351–389, http://dx.doi.org/10. 1007/11587392\_11.
- [10] C. MUSSO, N. OUDJANE, F. LE GLAND.*Improving regularized particle filters*, in "Sequential Monte Carlo Methods in Practice", New York, A. DOUCET, N. DE FREITAS, N. J. GORDON (editors), Statistics for Engineering and Information Science, Springer–Verlag, 2001, chap. 12, p. 247–271, http://dx.doi.org/10. 1007/978-1-4757-3437-9\_12.

# **Publications of the year**

## **Articles in International Peer-Reviewed Journal**

- [11] C. HERZET, A. DRÉMEAU, P. HÉAS. Model Reduction from Partial Observations, in "International Journal for Numerical Methods in Engineering", January 2018, vol. 113, nº 3, p. 479–511 [DOI: 10.1002/NME.5623], https://hal.archives-ouvertes.fr/hal-01394059.
- [12] P. HÉAS, C. HERZET.*Reduced Modeling of Unknown Trajectories*, in "Archives of Computational Methods in Engineering", January 2018, vol. 25, n<sup>o</sup> 1, p. 87-101, https://arxiv.org/abs/1702.08846 [DOI: 10.1007/s11831-017-9229-0], https://hal.archives-ouvertes.fr/hal-01490572.
- [13] M. LE CORVEC, C. JEZEQUEL, V. MONBET, N. FATIH, F. CHARPENTIER, H. TARIEL, C. BOUSSARD-PLÉDEL, B. BUREAU, O. LORÉAL, O. SIRE, E. BARDOU-JACQUET.*Mid-infrared spectroscopy of serum, a promising non-invasive method to assess prognosis in patients with ascites and cirrhosis*, in "PLoS ONE", October 2017, vol. 12, n<sup>o</sup> 10, e0185997 [*DOI :* 10.1371/JOURNAL.PONE.0185997], https://hal.archives-ouvertes.fr/hal-01625051.
- [14] V. MONBET, P. AILLIOT. Sparse vector Markov switching autoregressive models. Application to multivariate time series of temperature, in "Computational Statistics and Data Analysis", 2017, vol. 108, p. 40-51 [DOI: 10.1016/J.CSDA.2016.10.023], https://hal.archives-ouvertes.fr/hal-01250058.

#### **International Conferences with Proceedings**

- [15] P. HÉAS, C. HERZET. Optimal Low-Rank Dynamic Mode Decomposition, in "Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, New Orleans 2017", New Orleans, United States, March 2017, https://arxiv.org/abs/1701.01064 [DOI : 10.1109/ICASSP.2017.7952999], https://hal.archives-ouvertes.fr/hal-01429975.
- [16] S. SEN, A. CRINIÈRE, L. MEVEL, F. CÉROU, J. DUMOULIN. Seismic induced damage detection through parallel estimation of force and parameter using improved interacting Particle-Kalman filter, in "11th International Workshop on Structural Health Monitoring", San Francisco, United States, September 2017, https://hal.inria.fr/hal-01590713.
- [17] K. ZOUBERT-OUSSENI, C. VILLIEN, F. LE GLAND. Post-processing optimization of piecewise indoor trajectories based on IMU and RSS measurements, in "Proceedings of the 8th International Conference on Indoor Positioning and Indoor Navigation (IPIN)", Sapporo, Japan, IEEE, September 2017 [DOI: 10.1109/IPIN.2017.8115903], https://hal.inria.fr/hal-01671149.

#### **Conferences without Proceedings**

[18] S. SEN, A. CRINIÈRE, L. MEVEL, F. CÉROU, J. DUMOULIN. Estimation of time varying system parameters from ambient response using improved Particle-Kalman filter with correlated noise, in "EGU General Assembly 2017", Vienne, Austria, April 2017, https://hal.inria.fr/hal-01493840.

#### **Other Publications**

- [19] B. DELYON, F. CÉROU, A. GUYADER, M. ROUSSET.A Central Limit Theorem for Fleming-Viot Particle Systems with Hard Killing, October 2017, https://arxiv.org/abs/1709.06771 - 38 pages, https://hal.archivesouvertes.fr/hal-01614094.
- [20] P. HÉAS, C. HERZET. *Optimal Kernel-Based Dynamic Mode Decomposition*, 2017, https://arxiv.org/abs/1710. 10919 - working paper or preprint [*DOI* : 10.10919], https://hal.archives-ouvertes.fr/hal-01634542.

[21] M. ROUSSET, Y. XU, P.-A. ZITT.A Weak Overdamped Limit Theorem for Langevin Processes, September 2017, https://arxiv.org/abs/1709.09866 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01596954.

# **References in notes**

- [22] A. DOUCET, N. DE FREITAS, N. J. GORDON (editors). Sequential Monte Carlo methods in practice, Statistics for Engineering and Information Science, Springer–Verlag, New York, 2001, http://dx.doi.org/10.1007/978-1-4757-3437-9.
- [23] M. S. ARULAMPALAM, S. MAKSELL, N. J. GORDON, T. CLAPP.A tutorial on particle filters for online nonlinear / non-Gaussian Bayesian tracking, in "IEEE Transactions on Signal Processing", February 2002, vol. SP-50, n<sup>o</sup> 2 (Special issue on Monte Carlo Methods for Statistical Signal Processing), p. 174–188, http:// dx.doi.org/10.1109/78.978374.
- [24] M. BOCQUEL.Random finite sets in multi-target tracking: efficient sequential MCMC implementation, Department of Applied Mathematics, University of Twente, Enschede, October 2013, http://purl.utwente.nl/ publications/87496.
- [25] O. CAPPÉ, S. J. GODSILL, É. MOULINES. An overview of existing methods and recent advances in sequential Monte Carlo, in "Proceedings of the IEEE", May 2007, vol. 95, n<sup>o</sup> 5 (Special issue on Large–Scale Dynamic Systems), p. 899–924, http://dx.doi.org/10.1109/JPROC.2007.893250.
- [26] D. J. DALEY, J. GANI. *Epidemic modelling: An introduction*, Cambridge Studies in Mathematical Biology, Cambridge University Press, Cambridge, 1999, vol. 15, http://dx.doi.org/10.1017/CBO9780511608834.
- [27] P.–T. DE BOER, D. P. KROESE, S. MANNOR, R. Y. RUBINSTEIN. A tutorial on the cross-entropy method, in "Annals of Operations Research", January 2005, vol. 134 (Special issue on the Cross-Entropy Method for Combinatorial Optimization, Rare Event Simulation and Neural Computation), n<sup>o</sup> 1, p. 19–67, http://dx.doi. org/10.1007/s10479-005-5724-z.
- [28] P. DEL MORAL.*Mean field simulation for Monte Carlo integration*, Monographs on Statistics and Applied Probability, Chapman & Hall / CRC Press, London, 2013, vol. 126, http://www.crcpress.com/product/isbn/ 9781466504059.
- [29] P. DEL MORAL. Feynman-Kac formulae. Genealogical and interacting particle systems with applications, Probability and its Applications, Springer-Verlag, New York, 2004, http://dx.doi.org/10.1007/978-1-4684-9393-1.
- [30] P. DEL MORAL, L. MICLO.Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to nonlinear filtering, in "Séminaire de Probabilités XXXIV", Berlin, J. AZÉMA, M. ÉMERY, M. LEDOUX, M. YOR (editors), Lecture Notes in Mathematics, Springer-Verlag, 2000, vol. 1729, p. 1–145, http://dx.doi.org/10.1007/BFb0103798.
- [31] R. DOUC, É. MOULINES.Limit theorems for weighted samples with applications to sequential Monte Carlo methods, in "The Annals of Statistics", October 2008, vol. 36, n<sup>o</sup> 5, p. 2344–2376, http://dx.doi.org/10.1214/ 07-AOS514.

- [32] D. FOX, J. HIGHTOWER, L. LIAO, D. SCHULZ, G. BORRIELLO.*Bayesian filtering for location estimation*, in "IEEE Pervasive Computing", July/September 2003, vol. 2, n<sup>o</sup> 3, p. 24–33, http://dx.doi.org/10.1109/MPRV. 2003.1228524.
- [33] D. FOX, S. THRUN, W. BURGARD, F. DELLAERT. Particle filters for mobile robot localization, in "Sequential Monte Carlo Methods in Practice", New York, A. DOUCET, N. DE FREITAS, N. J. GORDON (editors), Statistics for Engineering and Information Science, Springer–Verlag, 2001, chap. 19, p. 401–428, http://dx. doi.org/10.1007/978-1-4757-3437-9\_19.
- [34] D. FRENKEL, B. SMIT. Understanding molecular simulation. From algorithms to applications, Computational Science Series, 2nd, Academic Press, San Diego, 2002, vol. 1, http://www.sciencedirect.com/science/book/ 9780122673511.
- [35] D. T. GILLESPIE. Approximate accelerated stochastic simulation of chemically reacting systems, in "The Journal of Chemical Physics", July 2001, vol. 115, n<sup>o</sup> 4, p. 1716–1733, http://dx.doi.org/10.1063/1.1378322.
- [36] P. GLASSERMAN.*Monte Carlo methods in financial engineering*, Applications of Mathematics, Springer–Verlag, New York, 2004, vol. 53, http://dx.doi.org/10.1007/978-0-387-21617-1.
- [37] P. GLASSERMAN, P. HEIDELBERGER, P. SHAHABUDDIN, T. ZAJIC. Multilevel splitting for estimating rare event probabilities, in "Operations Research", July–August 1999, vol. 47, n<sup>o</sup> 4, p. 585–600, http://dx.doi.org/ 10.1287/opre.47.4.585.
- [38] N. J. GORDON, D. J. SALMOND, A. F. M. SMITH. Novel approach to nonlinear / non-Gaussian Bayesian state estimation, in "IEE Proceedings, Part F", April 1993, vol. 140, n<sup>o</sup> 2, p. 107–113, http://dx.doi.org/10. 1049/ip-f-2.1993.0015.
- [39] F. GUSTAFSSON, F. GUNNARSSON, N. BERGMAN, U. FORSSELL, J. JANSSON, R. KARLSSON, P.-J. NORDLUND.*Particle filters for positioning, navigation, and tracking*, in "IEEE Transactions on Signal Processing", February 2002, vol. SP-50, n<sup>o</sup> 2 (Special issue on Monte Carlo Methods for Statistical Signal Processing), p. 425–437, http://dx.doi.org/10.1109/78.978396.
- [40] J. HOUSSINEAU. Representation and estimation of stochastic populations, School of Engineering and Physical Sciences, Heriot–Watt University, Edinburgh, August 2015, https://jeremiehoussineau.wordpress.com/ph-dthesis-all-versions/.
- [41] P. HÉAS, C. HERZET.Low-Rank Dynamic Mode Decomposition: Optimal Solution in Polynomial-Time, in "ArXiv e-prints", October 2016, https://arxiv.org/abs/1610.02962.
- [42] M. ISARD, A. BLAKE.CONDENSATION Conditional density propagation for visual tracking, in "International Journal of Computer Vision", August 1998, vol. 29, n<sup>o</sup> 1, p. 5–28, http://dx.doi.org/10.1023/ A:1008078328650.
- [43] G. KITAGAWA.*Monte Carlo filter and smoother for non–Gaussian nonlinear state space models*, in "Journal of Computational and Graphical Statistics", 1996, vol. 5, n<sup>o</sup> 1, p. 1–25, http://www.jstor.org/stable/1390750.
- [44] H. R. KÜNSCH.*Recursive Monte Carlo filters : Algorithms and theoretical analysis*, in "The Annals of Statistics", October 2005, vol. 33, n<sup>o</sup> 5, p. 1983–2021, http://www.jstor.org/stable/3448632.

- [45] P. L'ÉCUYER, V. DEMERS, B. TUFFIN. Rare events, splitting, and quasi-Monte Carlo, in "ACM Transactions on Modeling and Computer Simulation", April 2007, vol. 17, n<sup>o</sup> 2 (Special issue honoring Perwez Shahabuddin), Article 9, http://dx.doi.org/10.1145/1225275.1225280.
- [46] F. LE GLAND, N. OUDJANE.A robustification approach to stability and to uniform particle approximation of nonlinear filters : the example of pseudo-mixing signals, in "Stochastic Processes and their Applications", August 2003, vol. 106, n<sup>o</sup> 2, p. 279-316, http://dx.doi.org/10.1016/S0304-4149(03)00041-3.
- [47] F. LE GLAND, N. OUDJANE. Stability and uniform approximation of nonlinear filters using the Hilbert metric, and application to particle filters, in "The Annals of Applied Probability", February 2004, vol. 14, n<sup>o</sup> 1, p. 144–187, http://dx.doi.org/10.1214/aoap/1075828050.
- [48] J. S. LIU. Monte Carlo strategies in scientific computing, Springer Series in Statistics, Springer-Verlag, New York, 2001, http://dx.doi.org/10.1007/978-0-387-76371-2.
- [49] H. LOUVIN, E. DUMONTEIL, T. LELIÈVRE, M. ROUSSET, C. M. DIOP.Adaptive multilevel splitting for Monte Carlo particle transport, in "EPJ Web of Conferences", 2017, vol. 153 [DOI: 10.1051/EPJCONF/201715306006], https://hal.archives-ouvertes.fr/hal-01661012.
- [50] R. P. S. MAHLER. Multitarget Bayes filtering via first-order multitarget moments, in "IEEE Transactions on Aerospace and Electronic Systems", October 2003, vol. AES-39, n<sup>o</sup> 4, p. 1152–1178, http://dx.doi.org/10. 1109/TAES.2003.1261119.
- [51] B. RISTIĆ, M. S. ARULAMPALAM, N. J. GORDON. Beyond the Kalman Filter : Particle Filters for Tracking Applications, Artech House, Norwood, MA, 2004, http://www.artechhouse.com/international/books/945. aspx.
- [52] R. Y. RUBINSTEIN, D. P. KROESE. The cross-entropy method. A unified approach to combinatorial optimization, Monte Carlo simulation and machine learning, Information Science and Statistics, Springer–Verlag, New York, 2004, http://dx.doi.org/10.1007/978-1-4757-4321-0.
- [53] C. J. STONE. Consistent nonparametric regression (with discussion), in "The Annals of Statistics", July 1977, vol. 5, n<sup>o</sup> 4, p. 595–645, http://www.jstor.org/stable/2958783.
- [54] L. D. STONE, R. L. STREIT, T. L. CORWIN, K. L. BELL. *Bayesian multiple target tracking*, 2nd, Artech House, Norwood, MA, 2014, http://www.artechhouse.com/international/books/2090.aspx.
- [55] S. THRUN, W. BURGARD, D. FOX. Probabilistic robotics, Intelligent Robotics and Autonomous Agents, The MIT Press, Cambridge, MA, 2005, http://mitpress.mit.edu/books/probabilistic-robotics.
- [56] B.-N. VO, W.-K. MA. The Gaussian mixture probability hypothesis density filter, in "IEEE Transactions on Signal Processing", November 2006, vol. SP–54, n<sup>o</sup> 11, p. 4091–4104, http://dx.doi.org/10.1109/TSP.2006. 881190.
- [57] K. ZOUBERT-OUSSENI, C. VILLIEN, F. LE GLAND. Comparison of post-processing algorithms for indoor navigation trajectories, in "7th International Conference on Indoor Positioning and Indoor Navigation (IPIN)", Alcala de Henares, Spain, IEEE, October 2016 [DOI : 10.1109/IPIN.2016.7743590], https://hal.inria.fr/ hal-01423198.

[58] A. W. VAN DER VAART, J. A. WELLNER. *Weak convergence and empirical processes*, Springer Series in Statistics, Springer–Verlag, Berlin, 1996, http://dx.doi.org/10.1007/978-1-4757-2545-2.

# **Project-Team CAIRN**

# Energy Efficient Computing ArchItectures

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: CNRS Université Rennes 1 École normale supérieure de Rennes

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Architecture, Languages and Compilation

# **Table of contents**

1.	Personnel			
2.	Overall Objectives			
3.	. Research Program			
	3.1. Panorama	102		
	3.2. Reconfigurable Architecture Design	103		
	3.3. Compilation and Synthesis for Reconfigurable Platforms	103		
	3.4. Software Frameworks Developed by the Team	104		
4.	Application Domains			
5.	Highlights of the Year1			
6.	New Software and Platforms	106		
	6.1. Gecos	106		
	6.2. ID-Fix	107		
_	6.3. Platforms	107		
7.	New Results			
	7.1. Reconfigurable Architecture Design	108		
	7.1.1. Voltage Over-Scaling for Error-Resilient Applications	108		
	7.1.2. Stochastic Computation Elements with Correlated Input Streams	108		
	7.1.3. Fault Tolerant Architectures	109		
	7.1.4. Hardware Accelerated Simulation of Heterogeneous Platforms	109		
	7.1.5. Optical Interconnections for 3D Multiprocessor Architectures	109		
	7.1.6. Adaptive Dynamic Compilation for Low-Power Embedded Systems	110		
	7.1.7. Design Space Exploration for Iterative Stencil computations on FPGA accelerators	111		
	7.1.8. Energy-driven Accelerator Exploration for Heterogeneous Multiprocessor Architect	uresi i i		
	7.2. Compliation and Synthesis for Reconfigurable Platform	112		
	7.2.1. Superword-Level Parallelization Taskni gues for Time Critical Systems	112		
	7.2.2. Automatic Paranentzation Techniques for Time-Chucai Systems	112		
	7.2.5. Operator-Lever Approximate Computing	112		
	7.2.4. Dynamic Faut-Tolerant Mapping and Scheduling on Multi-core systems	113		
	7.2.5. Energy Constrained and Kear-Time Scheduling and Mapping on Mutheores 7.2.6. Real Time Scheduling of Reconfigurable Battery Powered Multi Core Platforms	113		
	7.2.0. Run-Time Management on Multicore Platforms	114		
8	Partnershins and Cooperations	114		
0.	8.1 National Initiatives	114		
	8.1.1 Labex CominLabs - 3DCORE (2014-2018)	114		
	8.1.2 Laber CominLabs - RELIASIC (2014-2018)	115		
	813 Laber CominLabs & Lebesgue - H-A-H (2014-2017)	115		
	8.14 Laber CominLabs - BBC (2016-2020)	115		
	8.1.5. Labex CominLabs - SHERPAM (2014-2018)	115		
	8.1.6. DGA RAPID - FLODAM (2017–2021)	116		
	8.2. European Initiatives	116		
	8.2.1. H2020 ARGO	116		
	8.2.2. ANR International ARTEFaCT	116		
	8.3. International Initiatives	117		
	8.3.1. Inria Associate Teams	117		
	8.3.2. Inria International Partners	117		
	8.3.2.1. LRS	117		
8.3.2.2. HARAMCOP				
	8.3.2.3. SPINACH	118		
	8.3.2.4. DARE	118		

	8.3	.2.5. Informal International Partners	118
	8.4. Inte	ernational Research Visitors	118
	8.4.1.	Visits of International Scientists	118
	8.4.2.	Visits to International Teams	118
	8.4.3.	Sabbatical programme	119
9.	Dissemina	ation	119
	9.1. Pro	omoting Scientific Activities	119
	9.1.1.	Chair of Conference Program Committees	119
	9.1.2.	Member of the Conference Program Committees	119
	9.1.3.	Member of the Editorial Boards of Journals	119
	9.1.4.	Invited Talks	119
	9.1.5.	Leadership within the Scientific Community	119
	9.1.6.	Scientific Expertise	120
	9.2. Tea	ching - Supervision - Juries	120
	9.2.1.	Teaching	120
	9.2.2.	Teaching Responsibilities	121
	9.2.3.	Supervision	121
10.	Bibliogra	aphy	122

# **Project-Team CAIRN**

Creation of the Project-Team: 2009 January 01

CAIRN is located on two campuses: Rennes (Beaulieu) and Lannion (ENSSAT).

# **Keywords:**

#### **Computer Science and Digital Science:**

- A1.1. Architectures
- A1.1.1. Multicore, Manycore
- A1.1.2. Hardware accelerators (GPGPU, FPGA, etc.)
- A1.1.8. Security of architectures
- A1.1.9. Fault tolerant systems
- A1.1.10. Reconfigurable architectures
- A1.1.12. Non-conventional architectures
- A1.2.5. Internet of things
- A1.2.6. Sensor networks
- A2.2. Compilation
- A2.2.1. Static analysis
- A2.2.4. Parallel architectures
- A2.2.5. GPGPU, FPGA, etc.
- A2.2.6. Adaptive compilation
- A4.4. Security of equipment and software
- A8.10. Computer arithmetic

## **Other Research Topics and Application Domains:**

- B4.5. Energy consumption
- B4.5.1. Green computing
- B4.5.2. Embedded sensors consumption
- B6.2.2. Radio technology
- B6.2.4. Optic technology
- B6.6. Embedded systems
- B8.1. Smart building/home
- B8.1.1. Energy for smart buildings
- B8.1.2. Sensor networks for smart buildings

# 1. Personnel

### **Research Scientists**

Olivier Sentieys [Team Leader, Senior Researcher (DR) Inria, HDR] François Charot [Inria, Researcher, Rennes] Tomofumi Yuki [Inria, Researcher, Rennes]

#### **Faculty Members**

Emmanuel Casseau [Professor, Univ. Rennes, ENSSAT, Lannion, HDR] Daniel Chillet [Professor, Univ. Rennes, ENSSAT, Lannion, HDR] Steven Derrien [Professor, Univ. Rennes, ISTIC, Rennes, HDR] Cédric Killian [Associate Professor, Univ. Rennes, IUT, Lannion] Angeliki Kritikakou [Associate Professor, Univ. Rennes, ISTIC, Rennes] Patrice Quinton [Ecole Normale Supérieure de Rennes, Emeritus, Rennes] Christophe Wolinski [Professor, Univ. Rennes, Director of ESIR, Rennes, HDR]

#### **Technical Staff**

Arnaud Carer [Research Engineer (half time), Univ. Rennes, Lannion] Pierre Guilloux [Univ. Rennes, Lannion] Pierre Halle [Inria, from Nov. 2017] Ali Hassan El Moussawi [Inria, Rennes] Mickael Dardaillon [Univ. Rennes, Rennes, from Dec. 2017] Thomas Lefeuvre [Univ. Rennes, Rennes, until Dec. 2017] Christophe Huriaux [Univ. Rennes, Rennes, until Dec. 2017]

#### **PhD Students**

Gabriel Gallin [CNRS, granted by CominLabs, from Oct. 2014] Jiating Luo [Univ. Rennes, granted by China Gov., from Nov. 2014] Van Dung Pham [Inria, granted by CominLabs, Lannion, from Dec. 2014] Aymen Gammoudi [Univ. Rennes, Lannion, from Sep. 2015] Rafail Psiakis [Univ. Rennes, MENRT grant, from Oct. 2015] Simon Rokicki [Univ. Rennes, granted by ENS Rennes, from Oct. 2015] Audrey Lucas [CNRS, granted by DGA-PEC, Lannion, from Jan. 2016] Genevieve Ndour [Univ. Rennes, granted by CEA Leti, Grenoble, from May 2016] Joel Ortiz Sosa [Inria, Lannion, from Oct. 2016] Nicolas Roux [Inria, granted by Brittany Region/LTC, Lannion, from Oct. 2016] Mael Gueguen [Univ. Rennes, MENRT grant, Rennes, from Nov. 2016] Minh Thanh Cong [Univ de Rennes, granted by USTH, Rennes, from May 2017] Thibaut Marty [Univ de Rennes, granted by H2020 ARGO and Brittany Region, Rennes, from Sep. 2017] Petr Dobias [Univ de Rennes, MENRT grant, Lannion, from Oct. 2017] Van Phu Ha [Inria, granted by ANR Artefact, Rennes, from Nov. 2017] Gaël Deest [Univ. Rennes I, MENRT grant, Rennes, until Jan. 2017] Rengarajan Ragavan [Univ. Rennes I, Lannion, until Jan. 2017] Xuan Chien Le [Inria, granted by Brittany Region/LTC, Lannion, until Mar. 2017] Baptiste Roux [Inria, granted by DGA and Inria, Rennes, until Sep. 2017] Benjamin Barrois [Univ. Rennes I, MENRT grant, Lannion, until Dec. 2017] Kleanthis Papachatzopoulos [Inria, Rennes, until Mar. 2017] Tara Petric [Inria, Rennes, until May 2017]

#### **Post-Doctoral Fellows**

Mansureh Shahraki Moghaddam [Inria, Rennes, from Dec 2017] Lei Mo [Inria, Rennes, from Apr 2017] Atef Dorai [Univ. Rennes, ATER, ENSSAT, Lannion, until Aug 2017] Imen Fassi [Univ. Rennes, Rennes, until Sep 2017] Ashraf El-Antably [Inria, Lannion, until May 2017] Imran Wali [Univ. Rennes, Lannion, until May 2017]

#### Administrative Assistants

Nadia Derouault [Assistant, Inria, Rennes] Emilie Carquin [Assistant, Univ. Rennes, ENSSAT, Lannion]

# 2. Overall Objectives

# 2.1. Overall Objectives

**Abstract** — The CAIRN project-team researches new architectures, algorithms and design methods for flexible, secure, fault-tolerant, and energy-efficient domain-specific system-on-chip (SoC). As performance and energy-efficiency requirements of SoCs, especially in the context of multi-core architectures, are continuously increasing, it becomes difficult for computing architectures to rely only on programmable processors solutions. To address this issue, we promote/advocate the use of reconfigurable hardware, i.e., hardware structures whose organization may change before or even during execution. Such reconfigurable chips offer high performance at a low energy cost, while preserving a high level of flexibility. The group studies these systems from three angles: (i) The invention and design of new reconfigurable architectures with an emphasis on flexible arithmetic operator design, dynamic reconfiguration management and low-power consumption. (ii) The development of their corresponding design flows (compilation and synthesis tools) to enable their automatic design from high-level specifications. (iii) The interaction between algorithms and architectures especially for our main application domains (wireless communications, wireless sensor networks and digital security).

**Keywords** — **Architectures:** Embedded Systems, System-on-Chip, Reconfigurable Architectures, Hardware Accelerators, Low-Power, Computer Arithmetic, Secure Hardware, Fault Tolerance. **Compilation and synthesis:** High-Level Synthesis, CAD Methods, Numerical Accuracy Analysis, Fixed-Point Arithmetic, Polyhedral Model, Constraint Programming, Source-to-Source Transformations, Domain-Specific Optimizing Compilers, Automatic Parallelization. **Applications:** Wireless (Body) Sensor Networks, High-Rate Optical Communications, Wireless Communications, Applied Cryptography.

The scientific goal of the CAIRN group is to research new hardware architectures for domain-specific SOCs, along with their associated design and compilation flows. We particularly focus on on-chip integration of specialized and reconfigurable accelerators. Reconfigurable architectures, whose hardware structure may be adjusted before or even during execution, originate from the possibilities opened up by Field Programmable Gate Arrays (FPGA) [57] and then by Coarse-Grain Reconfigurable Arrays (CGRA) [60], [72] [1]. Recent evolutions in technology and modern hardware systems confirm that reconfigurable systems are increasingly used in recent and future applications (see e.g. Intel/Altera or Xilinx/Zynq solutions). This architectural model has received a lot of attention in academia over the last two decades [63], and is now considered for industrial use in many application domains. One first reason is that the rapidly changing standards or applications require frequent device modifications. In many cases, software updates are not sufficient to keep devices on the market, while hardware redesigns remain too expensive. Second, the need to adapt the system to changing environments (e.g., wireless channel, harvested energy) is another incentive to use runtime dynamic reconfiguration. Moreover, with technologies at 28 nm and below, manufacturing problems strongly impact electrical parameters of transistors, and transient errors caused by particles or radiations also often appear during execution: error detection and correction mechanisms or autonomic self-control can benefit from reconfiguration capabilities.

As chip density increased, power or energy efficiency has become "the Grail" of all chip architects. With the end of Dennard scaling [67], multicore architectures are hitting the *utilisation wall* and the percentage of transistors in a chip that can switch at full frequency drops at a fast pace [61]. However, this unused portion of a chip also opens up new opportunities for computer architecture innovations. Building specialized processors or hardware accelerators can come with orders-of-magnitude gains in energy efficiency. Since from the beginning of CAIRN in 2009, we advocate the interest of heterogeneous multicores, in which general-purpose processors (GPPs) are integrated with specialized accelerators, especially when built on reconfigurable hardware, which provides the best trade-off between power, performance, cost and flexibility. During the period, it therefore turns out that the time has come for these heterogeneous manycore architectures.

Standard multicore architectures enable flexible software on fixed hardware, whereas reconfigurable architectures make possible **flexible software on flexible hardware**. However, designing reconfigurable systems poses several challenges: the definition of the architecture structure itself, along with its dynamic reconfiguration capabilities, and its corresponding compilation or synthesis tools. The scientific goal of CAIRN is therefore to leverage the background and past experience of its members to tackle these challenges. We propose to approach energy efficient reconfigurable architectures from three angles: (i) the invention and the design of new reconfigurable architectures or hardware accelerators, (ii) the development of their corresponding compilers and design methods, and (iii) the exploration of the interaction between applications and architectures.

# 3. Research Program

# 3.1. Panorama

The development of complex applications is traditionally split in three stages: a theoretical study of the algorithms, an analysis of the target architecture and the implementation. When facing new emerging applications such as high-performance, low-power and low-cost mobile communication systems or smart sensor-based systems, it is mandatory to strengthen the design flow by a joint study of both algorithmic and architectural issues.



Figure 1. CAIRN's general design flow and related research themes

Figure 1 shows the global design flow we propose to develop. This flow is organized in levels which refer to our three research themes: application optimization (new algorithms, fixed-point arithmetic, advanced representations of numbers), architecture optimization (reconfigurable and specialized hardware, application-specific processors, arithmetic operators and functions), and stepwise refinement and code generation (code transformations, hardware synthesis, compilation).

In the rest of this part, we briefly describe the challenges concerning **new reconfigurable platforms** in Section 3.2 and the issues on **compiler and synthesis tools** related to these platforms in Section 3.3.

# 3.2. Reconfigurable Architecture Design

Nowadays, FPGAs are not only suited for application specific algorithms, but also considered as fully-featured computing platforms, thanks to their ability to accelerate massively parallelizable algorithms much faster than their processor counterparts [75]. They also support to be dynamically reconfigured. At runtime, partially reconfigurable regions of the logic fabric can be reconfigured to implement a different task, which allows for a better resource usage and adaptation to the environment. Dynamically reconfigurable hardware can also cope with hardware errors by relocating some of its functionalities to another, sane, part of the logic fabric. It could also provide support for a multi-tasked computation flow where hardware tasks are loaded on-demand at runtime. Nevertheless, current design flows of FPGA vendors are still limited by the use of one partial bitstream for each reconfigurable region and for each design. These regions are defined at design time and it is not possible to use only one bitstream for multiple reconfigurable regions nor multiple chips. The multiplicity of such bitstreams leads to a significant increase in memory. Recent research has been conducted in the domain of task relocation on a reconfigurable fabric. All of the related work was conducted on architectures from commercial vendors (e.g., Xilinx, Altera) which share the same limitations: the inner details of the bitstream are not publicly known, which limits applicability of the techniques. To circumvent this issue, most dynamic reconfiguration techniques are either generating multiple bitstreams for each location [59] or implementing an online filter to relocate the tasks [69]. Both of these techniques still suffer from memory footprint and from the online complexity of task relocation.

Increasing the level and grain of reconfiguration is a solution to counterbalance the FPGA penalties. Coarsegrained reconfigurable architectures (CGRA) provide operator-level configurable functional blocks and wordlevel datapaths [76], [64], [74]. Compared to FPGA, they benefit from a massive reduction in configuration memory and configuration delay, as well as for routing and placement complexity. This in turns results in an improvement in the computation volume over energy cost ratio, although with a loss of flexibility compared to bit-level operations. Such constraints have been taken into account in the design of DART[7], Adres [72] or polymorphous computing fabrics[9]. These works have led to commercial products such as the PACT/XPP [58] or Montium from Recore systems, without however a real commercial success yet. Emerging platforms like Xilinx/Zynq or Intel/Altera are about to change the game.

In the context of emerging heterogenous multicore architecture, CAIRN advocates for associating generalpurpose processors (GPP), flexible network-on-chip and coarse-grain or fine-grain dynamically reconfigurable accelerators. We leverage our skills on microarchitecture, reconfigurable computing, arithmetic, and lowpower design, to discover and design such architectures with a focus on: -reduced energy per operation, improved application performance through acceleration, - hardware flexibility and self-adaptive behavior, tolerance to faults, computing errors, and process variation, - protections against side channel attacks, - limited silicon area overhead.

## 3.3. Compilation and Synthesis for Reconfigurable Platforms

In spite of their advantages, reconfigurable architectures, and more generally hardware accelerators, lack efficient and standardized compilation and design tools. As of today, this still makes the technology impractical for large-scale industrial use. Generating and optimizing the mapping from high-level specifications to reconfigurable hardware platforms are therefore key research issues, which have received considerable interest over the last years [62], [77], [73], [71], [70]. In the meantime, the complexity (and heterogeneity) of these platforms has also been increasing quite significantly, with complex heterogeneous multi-cores architectures becoming a *de facto* standard. As a consequence, the focus of designers is now geared toward optimizing overall system-level performance and efficiency [68]. Here again, existing tools are not well suited, as they fail at providing an unified programming view of the programmable and/or reconfigurable components implemented on the platform.

In this context, we have been pursuing our efforts to propose tools whose design principles are based on a tight coupling between the compiler and the target hardware architectures. We build on the expertise of the team members in High Level Synthesis (HLS) [4], ASIP optimizing compilers [10] and automatic parallelization for massively parallel specialized circuits [2]. We first study how to increase the efficiency of standard programmable processors by extending their instruction set to speed-up compute intensive kernels. Our focus is on efficient and exact algorithms for the identification, selection and scheduling of such instructions [5]. We address compilation challenges by borrowing techniques from high-level synthesis, optimizing compilers and automatic parallelization, especially when dealing with nested loop kernels. In addition, and independently of the scientific challenges mentioned above, proposing such flows also poses significant software engineering issues. As a consequence, we also study how leading edge software engineering techniques (Model Driven Engineering) can help the Computer Aided Design (CAD) and optimizing compiler communities prototyping new research ideas [3].

Efficient implementation of multimedia and signal processing applications (in software for DSP cores or as special-purpose hardware) often requires, for reasons related to cost, power consumption or silicon area constraints, the use of fixed-point arithmetic, whereas the algorithms are usually specified in floatingpoint arithmetic. Unfortunately, fixed-point conversion is very challenging and time-consuming, typically demanding up to 50% of the total design or implementation time. Thus, tools are required to automate this conversion. For hardware or software implementation, the aim is to optimize the fixed-point specification. The implementation cost is minimized under a numerical accuracy or an application performance constraint. For DSP-software implementation, methodologies have been proposed [6] to achieve fixed-point conversion. For hardware implementation, the best results are obtained when the word-length optimization process is coupled with the high-level synthesis [65]. Evaluating the effects of finite precision is one of the major and often the most time consuming step while performing fixed-point refinement. Indeed, in the word-length optimization process, the numerical accuracy is evaluated as soon as a new word-length is tested, thus, several times per iteration of the optimization process. Classical approaches are based on fixed-point simulations [66]. Leading to long evaluation times, they can hardly be used to explore the design space. Therefore, our aim is to propose closed-form expressions of errors due to fixed-point approximations that are used by a fast analytical framework for accuracy evaluation [8].

## **3.4. Software Frameworks Developed by the Team**

With the ever raising complexity of embedded applications and platforms, the need for efficient and customizable compilation flows is stronger than ever. This need of flexibility is even stronger when it comes to research compiler infrastructures that are necessary to gather quantitative evidence of the performance/energy or cost benefits obtained through the use of reconfigurable platforms. From a compiler point of view, the challenges exposed by these complex reconfigurable platforms are quite significant, since they require the compiler to extract and to expose an important amount of coarse and/or fine grain parallelism, to take complex resource constraints into consideration while providing efficient memory hierarchy and power management.

Because they are geared toward industrial use, production compiler infrastructures do not offer the level of flexibility and productivity that is required for compiler and CAD tool prototyping. To address this issue, we designed an extensible source-to-source compiler infrastructure that takes advantage of leading edge modeldriven object-oriented software engineering principles and technologies.

Figure 2 shows the global framework that is being developed in the group. Our compiler flow mixes several types of intermediate representations. The baseline representation is a simple tree-based model enriched with control flow information. This model is mainly used to support our source-to-source flow, and serves as the backbone for the infrastructure. We use the extensibility of the framework to provide more advanced representations along with their corresponding optimizations and code generation plug-ins. For example, for our pattern selection and accuracy estimation tools, we use a data dependence graph model in all basic blocks instead of the tree model. Similarly, to enable polyhedral based program transformations and analysis, we introduced a specific representation for affine control loops that we use to derive a Polyhedral Reduced Dependence Graph (PRDG). Our current flow assumes that the application is specified as a hierarchy of



Figure 2. CAIRN's general software development framework.

communicating tasks, where each task is expressed using C or Matlab/Scilab, and where the system-level representation and the target platform model are often defined using Domain Specific Languages (DSL).

**Gecos** (Generic Compiler Suite) is the main backbone of CAIRN's flow. It is an open source Eclipse-based flexible compiler infrastructure developed for fast prototyping of complex compiler passes. Gecos is a 100% Java based implementation and is based on modern software engineering practices such as Eclipse plugin or model-driven software engineering with EMF (Eclipse Modeling Framework). As of today, our flow offers the following features:

- An automatic floating-point to fixed-point conversion flow (for ASIC/FPGA and embedded processors). **ID.Fix** is an infrastructure for the automatic transformation of software code aiming at the conversion of floating-point data types into a fixed-point representation.
- A polyhedral-based loop transformation and parallelization engine (mostly targeted at HLS).
- A custom instruction extraction flow (for ASIP and dynamically reconfigurable architectures). **Durase** is developed for the compilation and the synthesis targeting reconfigurable platforms and the automatic synthesis of application specific processor extensions. It uses advanced technologies, such as graph matching together with constraint programming methods.
- Several back-ends to enable the generation of VHDL for specialized or reconfigurable IPs, and SystemC for simulation purposes (e.g., fixed-point simulations).

Gecos, ID.Fix or Durase have been demonstrated during "University Booths" in various conference such as IEEE/ACM DAC or DATE.

# 4. Application Domains

# 4.1. Panorama

**keywords:** Wireless (Body) Sensor Networks, High-Rate Optical Communications, Wireless Communications, Applied Cryptography.

Our research is based on realistic applications, in order to both discover the main needs created by these applications and to invent realistic and interesting solutions.

**Wireless Communication** is our privileged application domain. Our research includes the prototyping of (subsets of) such applications on reconfigurable and programmable platforms. For this application domain, the high computational complexity of the 5G Wireless Communication Systems calls for the design of high-performance and energy-efficient architectures. In **Wireless Sensor Networks** (WSN), where each wireless node is expected to operate without battery replacement for significant periods of time, energy consumption is the most important constraint. Sensor networks are a very dynamic domain of research due, on the one hand, to the opportunity to develop innovative applications that are linked to a specific environment, and on the other hand to the challenge of designing totally autonomous communicating objects.

Other important fields are also considered: hardware cryptographic and security modules, high-rate optical communications, machine learning, and multimedia processing.

# 5. Highlights of the Year

# 5.1. Highlights of the Year

Members of CAIRN published six papers accepted at IEEE/ACM Design Automation and Test in Europe for 2017, one of the major events in design automation.

[30] was among the few papers nominated for best paper at IEEE FPL.

# 6. New Software and Platforms

## **6.1. Gecos**

#### Generic Compiler Suite

KEYWORDS: Source-to-source compiler - Model-driven software engineering - Retargetable compilation SCIENTIFIC DESCRIPTION: The Gecos (Generic Compiler Suite) project is a source-to-source compiler infrastructure developed in the Cairn group since 2004. It was designed to enable fast prototyping of program analysis and transformation for hardware synthesis and retargetable compilation domains.

Gecos is Java based and takes advantage of modern model driven software engineering practices. It uses the Eclipse Modeling Framework (EMF) as an underlying infrastructure and takes benefits of its features to make it easily extensible. Gecos is open-source and is hosted on the Inria gforge.

The Gecos infrastructure is still under very active development, and serves as a backbone infrastructure to projects of the group. Part of the framework is jointly developed with Colorado State University and between 2012 and 2015 it was used in the context of the FP7 ALMA European project. The Gecos infrastructure is currently used by the EMMTRIX start-up, a spin-off from the ALMA project which aims at commercializing the results of the project, and in the context of the H2020 ARGO European project.

FUNCTIONAL DESCRIPTION: GeCoS provides a programme transformation toolbox facilitating parallelisation of applications for heterogeneous multiprocessor embedded platforms. In addition to targeting programmable processors, GeCoS can regenerate optimised code for High Level Synthesis tools.

- Participants: Tomofumi Yuki, Thomas Lefeuvre, Imèn Fassi, Mickael Dardaillon, Ali Hassan El Moussawi and Steven Derrien
- Partner: Université de Rennes 1
- Contact: Steven Derrien
- URL: http://gecos.gforge.inria.fr

# 6.2. ID-Fix

Infrastructure for the Design of Fixed-point systems

KEYWORDS: Energy efficiency - Dynamic range evaluation - Accuracy optimization - Fixed-point arithmetic - Analytic Evaluation - Embedded systems - Code optimisation

SCIENTIFIC DESCRIPTION: The different techniques proposed by the team for fixed-point conversion are implemented on the ID.Fix infrastructure. The application is described with a C code using floating-point data types and different pragmas, used to specify parameters (dynamic, input/output word-length, delay operations) for the fixed-point conversion. This tool determines and optimizes the fixed-point specification and then, generates a C code using fixed-point data types (ac\_fixed ) from Mentor Graphics. The infrastructure is made-up of two main modules corresponding to the fixed-point conversion (ID.Fix-Conv) and the accuracy evaluation (ID.Fix-Eval)

FUNCTIONAL DESCRIPTION: ID.Fix focuses on computational precision accuracy and can provide an optimised specification using fixed point arithmetic from a C source code with floating point data types. Fixed point arithmetic is very widely used in embedded systems as it provides better performance and is much more energy efficient. ID.Fix used an analytic programme model which means it can explore more solutions and thereby produce much more efficient code.

- Participant: Olivier Sentieys
- Partner: Université de Rennes 1
- Contact: Olivier Sentieys
- URL: http://idfix.gforge.inria.fr

# 6.3. Platforms

# 6.3.1. Zyggie

KEYWORDS: Health - Biomechanics - Wireless body sensor networks - Low power - Gesture recognition - Hardware platform - Software platform - Localization

SCIENTIFIC DESCRIPTION: Zyggie is a hardware and software wireless body sensor network platform. Each sensor node, attached to different parts of the human body, contains inertial sensors (IMU) (accelerometer, gyrometer, compass and barometer), an embedded processor and a low-power radio module to communicate data to a coordinator node connected to a computer, tablet or smartphone. One of the system's key innovations is that it collects data from sensors as well as on distances estimated from the power of the radio signal received to make the 3D location of the nodes more precise and thus prevent IMU sensor drift and power consumption overhead. Zyggie can be used to determine posture or gestures and mainly has applications in sport, healthcare and the multimedia industry.

FUNCTIONAL DESCRIPTION: The Zyggie sensor platform was developed to create an autonomous Wireless Body Sensor Network (WBSN) with the capabilities of monitoring body movements. The Zyggie platform is part of the BoWI project funded by CominLabs. Zyggie is composed of a processor, a radio transceiver and different sensors including an Inertial Measurement Unit (IMU) with 3-axis accelerometer, gyrometer, and magnetometer. Zyggie is used for evaluating data fusion algorithms, low power computing algorithms, wireless protocols, and body channel characterization in the BoWI project.

The Zyggie V2 prototype includes the following features: a 32-bit microcontroller to manage a custom MAC layer and processe quaternions based on IMU measures, and an UWB radio from DecaWave to measure distances between nodes with Time of Flight (ToF).

- Participants: Arnaud Carer and Olivier Sentieys
- Partners: Lab-STICC Université de Rennes 1
- Contact: Olivier Sentieys
- URL: http://www.bowi.cominlabs.ueb.eu/fr/zyggie-wbsn-platform



Figure 3. CAIRN's Ziggie platform for WBSN

# 7. New Results

# 7.1. Reconfigurable Architecture Design

## 7.1.1. Voltage Over-Scaling for Error-Resilient Applications

## Participants: Rengarajan Ragavan, Benjamin Barrois, Cédric Killian, Olivier Sentieys.

Voltage scaling has been used as a prominent technique to improve energy efficiency in digital systems, scaling down supply voltage effects in quadratic reduction in energy consumption of the system. Reducing supply voltage induces timing errors in the system that are corrected through additional error detection and correction circuits. In [43], we proposed voltage over-scaling based approximate operators for applications that can tolerate errors. We characterized the basic arithmetic operators using different operating triads (combination of supply voltage, body-biasing scheme and clock frequency) to generate models for approximate operators. Error-resilient applications can be mapped with the generated approximate operator models to achieve optimum trade-off between energy efficiency and error margin. Based on the dynamic speculation technique, best possible operating triad is chosen at runtime based on the user definable error tolerance margin of the application. In our experiments in 28nm FDSOI, we achieved maximum energy efficiency of 89% for basic operators like 8-bit and 16-bit adders at the cost of 20% Bit Error Rate (ratio of faulty bits over total bits) by operating them in near-threshold regime.

#### 7.1.2. Stochastic Computation Elements with Correlated Input Streams

Participants: Rengarajan Ragavan, Rahul Kumar Budhwani, Olivier Sentieys.

In recent years, shrinking size in integrated circuits has imposed a big challenge in maintaining the reliability in conventional computing. Stochastic Computing (SC) has been seen as a reliable, low-cost, and low-power alternative to overcome such issues. SC computes data in the form of bit streams of 1s and 0s. Therefore, SC outperforms conventional computing in terms of tolerance to soft error and uncertainty at the cost of increased computational time. Stochastic Computing with uncorrelated input streams requires streams to be highly independent for better accuracy. This results in more hardware consumption for conversion of binary numbers to stochastic streams. Correlation can be used to design Stochastic Computation Elements (SCE) with correlated input streams. These designs have higher accuracy and less hardware consumption. In [38], we proposed new SC designs to implement image processing algorithms with correlated input streams. Experimental results of proposed SC with correlated input streams show on average 37% improvement in accuracy with reduction of 50-90% in area and 20-85% in delay over existing stochastic designs.
#### 7.1.3. Fault Tolerant Architectures

Participants: Olivier Sentieys, Angeliki Kritikakou, Rafail Psiakis.

Error occurrence in embedded systems has significantly increased, whereas critical applications require reliable processors that combine performance with low cost and energy consumption. Very Long Instruction Word (VLIW) processors have inherent resource redundancy which is not constantly used due to application's fluctuating Instruction Level Parallelism (ILP). Approaches can benefit these additional resources to provide fault tolerance.

The reliability through idle slots utilization can be explored either at compile-time, increasing code size and storage requirements, or at run-time only inside the current instruction bundle, adding unnecessary time slots and degrading performance. To address this issue, we proposed a technique in [41] to explore the idle slots inside and across original and replicated instruction bundles reclaiming more efficiently the idle slots and creating a compact schedule. To achieve this, a dependency analysis is applied at run-time. The execution of both original and replicated instructions is allowed at any adequate function unit, providing higher flexibility on instruction scheduling. The proposed technique achieves up to 26% reduction in performance degradation over existing approaches.

When permanent and soft errors coexist, spare units have to be used or the executed program has to be modified through self-repair or by using several stored versions. However, these solutions introduce high area overhead for the additional resources, time overhead for the execution of the repair algorithm and storage overhead of the multi-versioning. To address these limitations, a hardware mechanism is proposed in [42] which at runtime replicates the instructions and schedules them at the idle slots considering the resource constraints. If a resource becomes faulty, the proposed approach efficiently rebinds both the original and replicated instructions during execution. In this way, the area overhead is reduced, as no spare resources are used, whereas time and storage overhead are not required. Results show up to 49% performance gain over existing techniques.

#### 7.1.4. Hardware Accelerated Simulation of Heterogeneous Platforms

Participants: Minh Thanh Cong, François Charot, Steven Derrien.

When considering designing heterogeneous multi-core platforms, the number of possible design combinations leads to a huge design space, with subtle trade-offs and design interactions. To reason about what design is best for a given target application requires detailed simulation of many different possible solutions. Simulation frameworks exist (such as gem5) and are commonly used to carry out these simulations. Unfortunately, these are purely software-based approaches and they do not allow a real exploration of the design space. Moreover, they do not really support highly heterogeneous multi-core architectures. These limitations motivate the study of the use of hardware to accelerate the simulation, and in particular of FPGA components. In this context, we are currently investigating the possibility of building hardware accelerated simulators using the HAsim simulation infrastructure, jointly developed by MIT and Intel. HAsim is an FPGA-accelerated simulator that is able to simulate a multicore with a high-detailed pipeline, cache hierarchy and detailed on-chip network on a single FPGA. A model of the RISC-V instruction set architecture suited to the HAsim infrastructure has been developed, its deployment on the Xeon+FPGA Intel platform is in progress. This work is done with the perspective of studying hardware accelerated simulation of heterogeneous multicore architectures mixing RISC-V cores and hardware accelerators.

#### 7.1.5. Optical Interconnections for 3D Multiprocessor Architectures

**Participants:** Jiating Luo, Ashraf El-Antably, Van Dung Pham, Cédric Killian, Daniel Chillet, Olivier Sentieys.

To address the issue of interconnection bottleneck in multiprocessor on a single chip, we study how an Optical Network-on-Chip (ONoC) can leverage 3D technology by stacking a specific photonics die. The objectives of this study target: i) the definition of a generic architecture including both electrical and optical components, ii) the interface between electrical and optical domains, iii) the definition of strategies (communication protocol) to manage this communication medium, and iv) new techniques to manage and reduce the power consumption of optical communications. The first point is required to ensure that electrical and optical components can be

used together to define a global architecture. Indeed, optical components are generally larger than electrical components, so a trade-off must be found between the size of optical and electrical parts. For the second point, we study how the interface can be designed to take applications needs into account. From the different possible interface designs, we extract a high-level performance model of optical communications from losses induced by all optical components to efficiently manage Laser parameters. Then, the third point concerns the definition of high-level mechanisms which can handle the allocation of the communication medium for each data transfer between tasks. This part consists in defining the protocol of wavelength allocation. Indeed, the optical wavelengths are a shared resource between all the electrical computing clusters and are allocated at run time according to application needs and quality of service. The last point concerns the definition of techniques allowing to reduce the power consumption of on-chip optical communications. The power of each Laser can be dynamically tuned in the optical/electrical interface at run time for a given targeted bit-error-rate. Due to the relatively high power consumption of such integrated Laser, we study how to define adequate policies able to adapt the laser power to the signal losses.

In [37] we designed an Optical-Network-Interface (ONI) to connect a cluster of several processors to the optical communication medium. This interface, constrained by the 10 Gb/s data-rate of the Lasers, integrates Error Correcting Codes (ECC) and a communication manager. This manager can select, at run-time, the communication mode to use depending on timing or power constraints. Indeed, as the use of ECC is based on redundant bits, it increases the transmission time, but saves power for a given Bit Error Rate (BER). Moreover, our ONI allows for data to be sent using several wavelengths in parallel, hence increasing transmission bandwidth. From the design of this interface, estimation in terms of power consumption and execution time have been obtained, as well as the energy per bit of each communication.

The optical medium can support multiple transactions at the same time on different wavelengths by using Wavelength Division Multiplexing (WDM). Moreover, multiple wavelengths can be gathered as highbandwidth channel to reduce transmission time. However, multiple signals sharing simultaneously a waveguide lead to inter-channel crosstalk noise. This problem impacts the Signal to Noise Ratio (SNR) of the optical signal, which increases the Bit Error Rate (BER) at the receiver side. In [39], we formulated the crosstalk noise and execution time models and then proposed a Wavelength Allocation (WA) method in a ring-based WDM ONoC to reach performance and energy trade-offs based on the application constraints. We showed that for a 16-core ONoC architecture using 12 wavelengths, more than  $10^5$  allocation solutions exist and only 51 are on a Pareto front giving a tradeoff between execution time and energy per bit (derived from the BER). These optimized solutions reduce the execution time by 37% or the energy from 7.6fJ/bit to 4.4fJ/bit.

We also proposed to explore the selection of laser power for each communication. This approach reduces the global power consumption by ensuring the targeted Bit Error Rate for each communication. To support laser power selection, we have also studied, designed and evaluated at transistor level different configurable laser drivers using a 28NM FDSOI technology.

#### 7.1.6. Adaptive Dynamic Compilation for Low-Power Embedded Systems

Participants: Steven Derrien, Simon Rokicki.

Single ISA-Heterogeneous multi-cores such as the ARM big.LITTLE have proven to be an attractive solution to explore different energy/performance trade-offs. Such architectures combine Out of Order cores with smaller in-order ones to offer different power/energy profiles. They however do not really exploit the characteristics of workloads (compute-intensive vs. control dominated).

In this work, we propose to enrich these architectures VLIW cores, which are very efficient at computeintensive kernels. To preserve the single ISA programming model, we resort to Dynamic Binary Translation as used in Transmeta Crusoe and NVidia Denver processors. Our proposed DBT framework targets the RISC-V ISA, for which both OoO and in-order implementations exist.

Since DBT operates at runtime, its execution time is directly perceptible by the user, hence severely constrained. As a matter of fact, this overhead has often been reported to have a huge impact on actual performance, and is considered as being the main weakness of DBT based solutions. This is particularly true when targeting a VLIW processor: the quality of the generated code depends on efficient scheduling; unfortunately scheduling is known to be the most time-consuming component of a JIT compiler or DBT. Improving the responsiveness of such DBT systems is therefore a key research challenge. This is however made very difficult by the lack of open research tools or platform to experiment with such platforms.

To address these issues, we have developed an open hardware/software platform supporting DBT. The platform was designed using HLS tools and validated on a FPGA board. The DBT uses RISC-V as host ISA, and can be retargeted to different VLIW configurations. Our platform uses custom hardware accelerators to improve the reactivity of our optimizing DBT flow. Our results [44] show that, compared to a software implementation, our approach offers speed-up by  $8 \times$  while consuming  $18 \times$  less energy.

Our current research work investigates how DBT techniques can be used to support runtime configurable VLIW cores. Such cores enable fine grain exploration of energy/performance trade-off by dynamically adjusting their number of execution slots, their register file size, etc.). More precisely, we build on our DBT framework to enable dynamic code specialization. Our first experimental results suggest that this approach leads to best-case performance and energy efficiency when compared against static VLIW configurations [54].

#### 7.1.7. Design Space Exploration for Iterative Stencil computations on FPGA accelerators

Participants: Steven Derrien, Gaël Deest, Tomofumi Yuki.

Iterative stencil computations arise in many application domains, ranging from medical imaging to numerical simulation. Since they are computationally demanding, a large body of work addressed the problem of parallelizing and optimizing stencils for multi-cores, GPUs, and FPGAs. Earlier attempts targeting FPGAs showed that the performance of such accelerators is the result of a complex interplay between the FPGA's raw computing power, the amount of on-chip memory it has, and the performance of the external memory system. They also illustrate how each application may have different requirements. For example, in the context of embedded vision, the designer's goal is often to find the design with minimum cost that matches realtime performance constraints (e.g., 4K@60fps). In an exascale context, the designer's goal is to maximize performance (measured in ops-per-second) for a given FPGA board, while maintaining power dissipation to a minimum. Based on these observations, we explore a family of design options that can accommodate a large set of requirements and constraints, by exposing trade-offs between computing power, bandwidth requirements, and FPGA resource usage. We have developed a code generator that produces HLS-optimized C/C++ descriptions of accelerator instances targeting emerging System on Chip platforms, (e.g., Xilinx Zynq or Intel SoC). Our family of designs builds upon the well-known tiling transformation, which we use to balance on-chip memory cost and off-chip bandwidth. To ease the exploration of this design space, we propose performance models to hone in on the most interesting design points, and show how they accurately lead to optimal designs. Our results demonstrate that the optimal choice depends on problem sizes and performance goals [30].

#### 7.1.8. Energy-driven Accelerator Exploration for Heterogeneous Multiprocessor Architectures Participants: Baptiste Roux, Olivier Sentieys.

Programming heterogeneous multiprocessor architectures combining multiple processor cores and hardware accelerators is a real challenge. Computer-aided design and development tools try to reduce the large design space by simplifying hardware software mapping mechanisms. However, energy consumption is not well supported in most of design space exploration methodologies due to the difficulty to fast and accurately estimate energy consumption. To this aim,we proposed and validated an exploration method for partitioning applications on software cores and hardware accelerators under energy-efficiency constraints. The methodology is based on energy and performance measurement of a tiny subset of the design space and an analytical formulation of the performance and energy of an application kernel mapped on a heterogeneous architecture. This closed-form expression is captured and solved using Mixed Integer Linear Programming, which allows for very fast exploration resulting in the optimal solution. The approach is validated on two applications kernels using Zynq-based architecture showing more than 12% acceleration speed-up and energy saving compared to standard approaches. Results also show that the most energy-efficient solution is application- and platform-dependent and moreover hardly predictable, which highlights the need for fast exploration.

#### 7.2. Compilation and Synthesis for Reconfigurable Platform

#### 7.2.1. Superword-Level Parallelism-Aware Word Length Optimization

Participants: Steven Derrien, Ali Hassan El Moussawi.

Many embedded processors do not support floating-point arithmetic in order to comply with strict cost and power consumption constraints. But, they generally provide support for SIMD as a mean to improve performance for little cost overhead. Achieving good performance when targeting such processors requires the use of fixed-point arithmetic and efficient exploitation of SIMD data-path. To reduce time-to-market, automatic SIMDization – such as superword level parallelism (SLP) extraction – and floating-point to fixedpoint conversion methodologies have been proposed. In [33], we showed that applying these transformations independently is not efficient. We proposed an SLP-aware word length optimization algorithm to jointly perform floating-point to fixed-point conversion and SLP extraction. We implemented the proposed approach in a source-to-source compiler framework and evaluated it on several embedded processors. Experimental results illustrated the validity of our approach with performance improvement by up to 40% for a limited loss in accuracy.

#### 7.2.2. Automatic Parallelization Techniques for Time-Critical Systems

Participants: Steven Derrien, Imen Fassi, Thomas Lefeuvre.

Real-time systems are ubiquitous, and many of them play an important role in our daily life. In hard real-time systems, computing the correct results is not the only requirement. In addition, the results must be produced within pre-determined timing constraints, typically deadlines. To obtain strong guarantees on the system temporal behavior, designers must compute upper bounds of the Worst-Case Execution Times (WCET) of the tasks composing the system. WCET analysis is confronted with two challenges: (i) extracting knowledge of the execution flow of an application from its machine code, and (ii) modeling the temporal behavior of the target platform. Multi-core platforms make the latter issue even more challenging, as interference caused by concurrent accesses to shared resources have also to be modeled. Accurate WCET analysis is facilitated by *predictable* hardware architectures. For example, platforms using ScratchPad Memories (SPMs) instead of caches are considered as more predictable. However SPM management is left to the programmer-managed, making them very difficult to use, especially when combined with complex loop transformations needed to enable task level parallelization. Many researches have studied how to combine automatic SPM management with loop parallelization at the compiler level. It has been shown that impressive average-case performance improvements could be obtained on compute intensive kernels, but their ability to reduce WCET estimates remains to be demonstrated, as the transformed code does not lends itself well to WCET analysis.

In the context of the ARGO project, and in collaboration with members of the PACAP team, we have studied how parallelizing compilers techniques should be revisited in order to help WCET analysis tools. More precisely, we have demonstrated the ability of polyhedral optimization techniques to reduce WCET estimates in the case of sequential codes, with a focus on locality improvement and array contraction. We have shown on representative real-time image processing use cases that they could bring significant improvements of WCET estimates (up to 40%) provided that the WCET analysis process is guided with automatically generated flow annotations [31].

#### 7.2.3. Operator-Level Approximate Computing

Participants: Benjamin Barrois, Olivier Sentieys.

Many applications are error-resilient, allowing for the introduction of approximations in the calculations, as long as a certain accuracy target is met. Traditionally, fixed-point arithmetic is used to relax accuracy, by optimizing the bit-width. This arithmetic leads to important benefits in terms of delay, power and area. Lately, several hardware approximate operators were invented, seeking the same performance benefits. However, a fair comparison between the usage of this new class of operators and classical fixed-point arithmetic with careful truncation or rounding, has never been performed. In [27], we first compare approximate and fixed-point arithmetic operators in terms of power, area and delay, as well as in terms of induced error, using many

state-of-the-art metrics and by emphasizing the issue of data sizing. To perform this analysis, we developed a design exploration framework, *ApxPerf*, which guarantees that all operators are compared using the same operating conditions. Moreover, operators are compared in several classical real-life applications leveraging relevant metrics. In [27], we show that considering a large set of parameters, existing approximate adders and multipliers tend to be dominated by truncated or rounded fixed-point ones. For a given accuracy level and when considering the whole computation data-path, fixed-point operators are several orders of magnitude more accurate while spending less energy to execute the application. A conclusion of this study is that the entropy of careful sizing is always lower than approximate operators, since it require significantly less bits to be processed in the data-path and stored. Approximated data therefore always contain on average a greater amount of costly erroneous, useless information.

In [26] we performed a comparison between custom fixed-point (FxP) and floating-point (FIP) arithmetic, applied to bidimensional K-means clustering algorithm. First, FxP and FIP arithmetic operators are compared in terms of area, delay and energy, for different bitwidth, using the *ApxPerf2.0* framework. Finally, both are compared in the context of K-means clustering. The direct comparison shows the large difference between 8-to-16-bit FxP and FIP operators, FIP adders consuming  $5-12\times$  more energy than FxP adders, and multipliers  $2-10\times$  more. However, when applied to K-means clustering algorithm, the gap between FxP and FIP tightens. Indeed, the accuracy improvements brought by FIP make the computation more accurate and lead to an accuracy equivalent to FxP with less iterations of the algorithm, proportionally reducing the global energy spent. The 8-bit version of the algorithm becomes more profitable using FIP, which is 80% more accurate with only  $1.6\times$  more energy.

#### 7.2.4. Dynamic Fault-Tolerant Mapping and Scheduling on Multi-core systems

Participants: Emmanuel Casseau, Petr Dobias.

Demand on multi-processor systems for high performance and low energy consumption still increases in order to satisfy our requirements to perform more and more complex computations. Moreover, the transistor size gets smaller and their operating voltage is lower, which goes hand in glove with higher susceptibility to system failure. In order to ensure system functionality, it is necessary to conceive fault-tolerant systems. One way to tackle this issue is to makes use of both the redundancy and reconfigurable computing, especially when multi-processor platforms are targeted. Actually, multi-processor platforms can be less vulnerable when one processor is faulty because other processors can take over its scheduled tasks.

In this context, we investigate how to dynamically map and schedule tasks onto homogeneous faulty processors. We developed a run-time algorithm based on the primary/backup approach which is commonly used for its minimal resources utilization and high reliability. Its principal rule is that, when a task arrives, the system creates two identical copies: the primary copy and the backup copy. Several policies have been studied and their performances have been analyzed. We are currently refining the algorithm to reduce its complexity without decreasing performance. This work is done in collaboration with Oliver Sinnen, PARC Lab., the University of Auckland.

#### 7.2.5. Energy Constrained and Real-Time Scheduling and Mapping on Multicores

Participants: Olivier Sentieys, Angeliki Kritikakou, Lei Mo.

Multicore architectures are now widely used in energy-constrained real-time systems, such as energyharvesting wireless sensor networks. To take advantage of these multicores, there is a strong need to balance system energy, performance and Quality-of-Service (QoS). The Imprecise Computation (IC) model splits a task into mandatory and optional parts allowing to tradeoff QoS. We focus on the problem of mapping, i.e. allocating and scheduling, IC-tasks to a set of processors to maximize system QoS under real-time and energy constraints, which we formulate as a Mixed Integer Linear Programming (MILP) problem. However, state-of-the-art solving techniques either demand high complexity or can only achieve feasible (suboptimal) solutions. We develop an effective decomposition-based approach in [40] to achieve an optimal solution while reducing computational complexity. It decomposes the original problem into two smaller easier-tosolve problems: a master problem for IC-tasks allocation and a slave problem for IC-tasks scheduling. We also provide comprehensive optimality analysis for the proposed method. Through the simulations, we validate and demonstrate the performance of the proposed method, resulting in an average 55% QoS improvement with regards to published techniques.

#### 7.2.6. Real-Time Scheduling of Reconfigurable Battery-Powered Multi-Core Platforms

Participants: Daniel Chillet, Aymen Gammoudi.

Reconfigurable real-time embedded systems are constantly increasingly used in applications like autonomous robots or sensor networks. Since they are powered by batteries, these systems have to be energy-aware, to adapt to their environment and to satisfy real-time constraints. For energy harvesting systems, regular recharges of battery can be estimated, and by including this parameter in the operating system, it is then possible to develop strategy able to ensure the best execution of the application until the next recharge. In this context, operating system services must control the execution of tasks to meet the application constraints. Our objective concerns the proposition of a new real-time scheduling strategy that considers execution constraints such as the deadline of tasks and the energy for heterogeneous architectures. For such systems, we first addressed homogeneous architectures and extended our work for heterogeneous systems for which each task has different execution parameters. For these two architectures models, we formulated the problem as an ILP optimisation problem that can be solved by classical solvers. Assuming that the energy consumed by the communication is dependent on the distance between processors, we proposed a mapping strategy to minimise the total cost of communication between processors by placing the dependent tasks as close as possible to each other. The proposed strategy guarantees that, when a task is mapped into the system and accepted, it is then correctly executed prior to the task deadline. Finally, as on-line scheduling is targeted for this work, we proposed heuristics to solve these problems in efficient way. These heuristics are based on the previous packing strategy developed for the mono-processor architecture case.

#### 7.2.7. Run-Time Management on Multicore Platforms

Participant: Angeliki Kritikakou.

In real-time mixed-critical systems, Worst-Case Execution Time analysis (WCET) is required to guarantee that timing constraints are respected —at least for high criticality tasks. However, the WCET is pessimistic compared to the real execution time, especially for multicore platforms. As WCET computation considers the worst-case scenario, it means that whenever a high criticality task accesses a shared resource in multicore platforms, it is considered that all cores use the same resource concurrently. This pessimism in WCET computation leads to a dramatic under utilization of the platform resources, or even failing to meet the timing constraints. In order to increase resource utilization while guaranteeing real-time guarantees for high criticality tasks, previous works proposed a run-time control system to monitor and decide when the interferences from low criticality tasks cannot be further tolerated. However, in the initial approaches, the points where the controller is executed were statically predefined. We propose a dynamic run-time control in [19] which adapts its observations to on-line temporal properties, increasing further the dynamism of the approach, and mitigating the unnecessary overhead implied by existing static approaches. Our dynamic adaptive approach allows to control the ongoing execution of tasks based on run-time information, and increases further the gains in terms of resource utilization compared with static approaches.

# 8. Partnerships and Cooperations

#### 8.1. National Initiatives

#### 8.1.1. Labex CominLabs - 3DCORE (2014-2018)

**Participants:** Olivier Sentieys, Daniel Chillet, Cédric Killian, Jiating Luo, Van Dung Pham, Ashraf El-Antably.

3DCORE (3D Many-Core Architectures based on Optical Network on Chip) is a project investigating new solutions based on silicon photonics to enhance by 2 to 3 magnitude orders energy efficiency and data rate

of on-chip interconnect in the context of a many-core architecture. Moreover, 3DCore will take advantage of 3D technologies to design a specific optical layer suitable for a flexible and energy efficient high-speed optical network on chip (ONoC). 3DCORE involves CAIRN, FOTON (Rennes, Lannion) and Institut des Nanotechnologies de Lyon. For more details see http://www.3d-opt-many-cores.cominlabs.ueb.eu.

#### 8.1.2. Labex CominLabs - RELIASIC (2014-2018)

Participants: Emmanuel Casseau, Imran Wali.

RELIASIC (Reliable Asic) will address the issue of fault-tolerant computation with a bottom-up approach, starting from an existing application as a use case (a GPS receiver) and adding some redundant mechanisms to allow the GPS receiver to be tolerant to transient errors due to low voltage supply. RELIASIC involves CAIRN, Lab-STICC (Lorient) and IETR (Rennes, Nantes). For more details see http://www.reliasic.cominlabs.ueb.eu In this project, CAIRN is in charge of the analysis and design of arithmetic operators for fault tolerance. We focus on the hardware implementations of conventional arithmetic operators such as adders, multipliers. We also propose a lightweight design and assessment framework for arithmetic operators with reduced-precision redundancy.

#### 8.1.3. Labex CominLabs & Lebesgue - H-A-H (2014-2017)

Participants: Arnaud Tisserand, Gabriel Gallin, Audrey Lucas.

H-A-H for *Hardware and Arithmetic for Hyperelliptic Curves Cryptography* is a project on advanced arithmetic representation and algorithms for hyper-elliptic curve cryptography. It will provide novel implementations of HECC based cryptographic algorithms on custom hardware platforms. H-A-H involves CAIRN (Lannion) and IRMAR (Rennes). For more details see http://h-a-h.inria.fr/.

#### 8.1.4. Labex CominLabs - BBC (2016-2020)

Participants: Olivier Sentieys, Cédric Killian, Joel Ortiz Sosa.

The aim of the BBC (on-chip wireless Broadcast-Based parallel Computing) project is to evaluate the use of wireless links between cores inside chips and to define new paradigms. Using wireless communications enables broadcast capabilities for Wireless Networks on Chip (WiNoC) and new management techniques for memory hierarchy and parallelism. The key objectives concern improvement of power consumption, estimation of achievable data rates, flexibility and reconfigurability, size reduction and memory hierarchy management. For more details see http://www.bbc.cominlabs.ueb.eu In this project, CAIRN will address new low-power MAC (media access control) technique based on CDMA access as well as broadcast-based fast cooperation protocol designed for resource sharing (bandwidth, distributed memory, cache coherency) and parallel programming.

#### 8.1.5. Labex CominLabs - SHERPAM (2014-2018)

#### Participant: Patrice Quinton.

Heart failure and peripheral artery disease patients require early detection of health problems in order to prevent major risk of morbidity and mortality. Evidence shows that people recover from illness or cope with a chronic condition better if they are in a familiar environment (i.e., at home) and if they are physically active (i.e., practice sports). The goal of the Sherpam project is to design, implement, and validate experimentally a monitoring system allowing biophysical data of mobile subjects to be gathered and exploited in a continuous flow. Transmission technologies available to mobile users have been improved a lot during the last two decades, and such technologies offer interesting prospects for monitoring the health of people anytime and anywhere. The originality of the Sherpam project is to rely simultaneously and in an agile way on several kinds of wireless networks in order to ensure the transmission of biometric data, while coping with network disruptions. Sherpam also develops new signal processing algorithms for activity quantification and recognition which represent now a major social and public health issue (monitoring of elderly patient, personalized quantification activity, etc.). Sherpam involves research teams from several scientific domains and from several laboratories of Brittany (IRISA/CASA, LTSI, M2S, CIC-IT 1414-CHU Rennes and LAUREPS). For more details see http://www.sherpam.cominlabs.ueb.eu

#### 8.1.6. DGA RAPID - FLODAM (2017-2021)

Participants: Olivier Sentieys, Angeliki Kritikakou.

FLODAM is an industrial research project for methodologies and tools dedicated to the hardening of embedded multi-core processor architectures. The goal is to: 1) evaluate the impact of the natural or artificial environments on the resistance of the system components to faults based on models that reflect the reality of the system environment, 2) the exploration of architecture solutions to make the multi-core architectures fault tolerant to transient or permanent faults and 3) test and evaluate the proposed fault tolerant architecture solutions and compare the results under different scenarios provided by the fault models.

#### 8.2. European Initiatives

#### 8.2.1. H2020 ARGO

Participants: Steven Derrien, Olivier Sentieys, Imen Fassi, Ali Hassan El Moussawi.

Program: H2020-ICT-04-2015 Project acronym: ARGO Project title: WCET-Aware Parallelization of Model-Based Applications for Heterogeneous Parallel Systems Duration: Feb. 2016 - Feb. 2019 Coordinator: KIT Other partners: KIT (DE), UR1/Inria/CAIRN (FR), Recore Systems (NL), TEI-WG (GR), Scilab Ent. (FR), Absint (DE), DLR (DE), Fraunhofer (DE)

Increasing performance and reducing cost, while maintaining safety levels and programmability are the key demands for embedded and cyber-physical systems, e.g. aerospace, automation, and automotive. For many applications, the necessary performance with low energy consumption can only be provided by customized computing platforms based on heterogeneous many-core architectures. However, their parallel programming with time-critical embedded applications suffers from a complex toolchain and programming process. ARGO will address this challenge with a holistic approach for programming heterogeneous multi- and many-core architectures using automatic parallelization of model-based real-time applications. ARGO will enhance WCET-aware automatic parallelization by a cross-layer programming approach combining automatic tool-based and user-guided parallelization to reduce the need for expertise in programming parallel heterogeneous architectures. The ARGO approach will be assessed and demonstrated by prototyping comprehensive time-critical applications from both aerospace and industrial automation domains on customized heterogeneous many-core platforms.

#### 8.2.2. ANR International ARTEFaCT

Participants: Olivier Sentieys, Benjamin Barrois, Tara Petric, Tomofumi Yuki.

Program: ANR International France-Switzerland Project acronym: ARTEFaCT Project title: AppRoximaTivE Flexible Circuits and Computing for IoT Duration: Feb. 2016 - Dec. 2019 Coordinator: CEA Other partners: CEA-LETI (FR), CAIRN (FR), EPFL (SW)

The ARTEFaCT project aims to build on the preliminary results on inexact and exact near-threshold and sub-threshold circuit design to achieve major energy consumption reductions by enabling adaptive accuracy control of applications. ARTEFaCT proposes to address, in a consistent fashion, the entire design stack, from physical hardware design, up to software application analysis, compiler optimizations, and dynamic energy management. We do believe that combining sub-near-threshold with inexact circuits on the hardware side and, in addition, extending this with intelligent and adaptive power management on the software side will produce outstanding results in terms of energy reduction, i.e., at least one order of magnitude, in IoT applications. The project will contribute along three research directions: (1) approximate, ultra low-power circuit design, (2) modeling and analysis of variable levels of computation precision in applications, and (3) accuracy-energy trade- offs in software.

# 8.3. International Initiatives

#### 8.3.1. Inria Associate Teams

8.3.1.1. IoTA

Title: Ultra-Low Power Computing Platform for IoT leveraging Controlled Approximation International Partner (Institution - Laboratory - Researcher):

Ecole Polytechnique Fédérale de Lausanne (Switzerland) - Christian Enz

Start year: 2017

See also: https://team.inria.fr/cairn/IOTA

Energy issues are central to the evolution of the Internet of Things (IoT), and more generally to the ICT industry. Current low-power design techniques cannot support the estimated growth in number of IoT objects and at the same time keep the energy consumption within sustainable bounds, both on the IoT node side and on cloud/edge-cloud side. This project aims to build on the preliminary results on inexact and exact sub/near-threshold circuit design to achieve major energy consumption reductions by enabling adaptive accuracy control of applications. Advanced ultra lowpower hardware design methods utilize very low supply voltage, such as in near-threshold and sub-threshold designs. These emerging technologies are very promising avenues to decrease active and stand-by-power in electronic devices. To move another step forward, recently, approximate computing has become a major field of research in the past few years. IoTA proposes to address, in a consistent fashion, the entire design stack, from hardware design, up to software application analysis, compiler optimizations, and dynamic energy management. We do believe that combining sub-near-threshold with inexact circuits on the hardware side and, in addition, extending this with intelligent and adaptive power management on the software side will produce outstanding results in terms of energy reduction, i.e., at least one order of magnitude, in IoT. The main scientific challenge is twofold: (1) to add adaptive accuracy to hardware blocks built in near/sub threshold technology and (2) to provide the tools and methods to program and make efficient use of these hardware blocks for applications in the IoT domain. This entails developing approximate computing units, on one side, and methods and tools, on the other side, to rigorously explore trade-offs between accuracy and energy consumption in IoT systems. The expertise of the members of the two teams is complementary and covers all required technical knowledge necessary to reach our objectives, i.e., ultra low power hardware design (EPFL), approximate operators and functions (Inria, EPFL), formal analysis of precision in algorithms (Inria), and static and dynamic energy management (Inria, EPFL). Finally, the proof of concept will consist of results on (1) an adaptive, inexact or exact, ultra-low power microprocessor in 28 nm process and (2) a real prototype implemented in an FPGA platform combining processors and hardware accelerators. Several software use-cases relevant for the IoT domain will be considered, e.g., embedded vision, IoT sensors data fusion, to practically demonstrate the benefits of our approach.

#### 8.3.2. Inria International Partners

8.3.2.1. LRS

Title: Loop unRolling Stones: compiling in the polyhedral model

International Partner (Institution - Laboratory - Researcher):

Colorado State University (United States) - Department of Computer Science - Prof. Sanjay Rajopadhye

#### 8.3.2.2. HARAMCOP

Title: Hardware accelerators modeling using constraint-based programming

International Partner (Institution - Laboratory - Researcher):

Lund University (Sweden) - Department of Computer Science - Prof. Krzysztof Kuchcinski

#### 8.3.2.3. SPINACH

Title: Secure and low-Power sensor Networks Circuits for Healthcare embedded applications

International Partner (Institution - Laboratory - Researcher):

University College Cork (Ireland) - Department of Electrical and Electronic Engineering -Prof. Liam Marnane and Prof. Emanuel Popovici

Arithmetic operators for cryptography, side channel attacks for security evaluation, energyharvesting sensor networks, and sensor networks for health monitoring.

#### 8.3.2.4. DARE

Title: Design space exploration Approaches for Reliable Embedded systems

International Partner (Institution - Laboratory - Researcher):

IMEC (Belgium) - Francky Catthoor

Methodologies to design low cost and efficient techniques for safety-critical embedded systems, Design Space Exploration (DSE), run-time dynamic control mechanisms.

#### 8.3.2.5. Informal International Partners

LSSI laboratory, Québec University in Trois-Rivières (Canada), Design of architectures for digital filters and mobile communications.

Department of Electrical and Computer Engineering, University of Patras (Greece), Wireless Sensor Networks, Worst-Case Execution Time, Priority Scheduling.

Karlsruhe Institute of Technology - KIT (Germany), Loop parallelization and compilation techniques for embedded multicores.

Ruhr - University of Bochum - RUB (Germany), Reconfigurable architectures.

University of Science and Technology of Hanoi (Vietnam), Participation of several CAIRN's members in the Master ICT / Embedded Systems.

#### **8.4. International Research Visitors**

#### 8.4.1. Visits of International Scientists

Mattia Cacciotti, Ecole Polytechnique Fédérale de Lausanne (Switzerland), from May 2017 until June 2017.

Emna Hammami, University of Tunis, from April 2017 until June 2017.

Prof. Stanislaw Piestrak, Univ de Lorraine, June 2017.

#### 8.4.2. Visits to International Teams

P. Quinton was invited in Passau University (Passau, Germany) by Prof. Chris Lengauer during one week in June 2017, and gave an invited seminar on the synthesis of parallel architectures.

P. Quinton was invited by Prof. Daniel Massicotte of Université de Trois-Rivières (Québec) in October 2017 to cooperate on the design of FPGA hardware accelerators for electric simulation. His stay was supported by a grant of the RESMIQ (regroupement stratégique en microsystèmes du Québec). He gave an invited seminar on the synthesis of data-flow parallel systems.

#### 8.4.3. Sabbatical programme

#### Casseau Emmanuel

Date: Aug 2016 - Jul 2017

Institution: University of Auckland (New Zealand), Parallel and Reconfigurable Research Lab. of the Electrical and Computer Engineering department.

The goal of the project was to propose dynamic mapping and scheduling algorithms dedicated to unreliable heterogeneous platforms, enabling self-adaptive and resource-aware computing.

# 9. Dissemination

#### 9.1. Promoting Scientific Activities

#### 9.1.1. Chair of Conference Program Committees

• O. Sentieys was Track Chair at IEEE NEWCAS.

#### 9.1.2. Member of the Conference Program Committees

- D. Chillet was member of the technical program committee of HiPEAC RAPIDO, HiPEAC WRC, MCSoC, DCIS, ComPAS, DASIP, LP-EMS, ARC.
- S. Derrien was a member of technical program committee of IEEE FPL and ARC conferences and of WRC and Impact workshops.
- O. Sentieys was a member of technical program committee of IEEE/ACM DATE, IEEE FPL, ACM ENSSys, ACM SBCCI, IEEE ReConFig, FPGA4GPC.

#### 9.1.3. Member of the Editorial Boards of Journals

- D. Chillet is member of the Editor Board of Journal of Real-Time Image Processing (JRTIP).
- O. Sentieys is member of the editorial board of Journal of Low Power Electronics and International Journal of Distributed Sensor Networks.

#### 9.1.4. Invited Talks

- O. Sentieys gave an invited talk at FETCH (École d'hiver Francophone sur les Technologies de Conception des Systèmes embarqués Hétérogènes), Mont Tremblant, Canada, January 2017 on "Need more Energy Efficiency? Agree to Compute Inexactly".
- O. Sentieys gave an invited talk at GDR SoC<sup>2</sup>, Paris, France, November 2017 on "Controlling Inexact Computations at Compile Time and Runtime".
- O. Sentieys gave an invited talk at IoT2Sustain Workshop, London, UK, July 2017 on "Challenges in Energy Efficiency of Computing Architectures: from Sensors to Clouds".
- O. Sentieys gave an invited course at ARCHI Spring School, Nancy, France, March 2017 on "Design of VLSI Integrated Circuits A (very) deep dive into processors".

#### 9.1.5. Leadership within the Scientific Community

- D. Chillet is member of the Board of Directors of Gretsi Association.
- D. Chillet is co-animator of the topics "Connected Objects" and "Near Sensor Computing" of GDR SoC<sup>2</sup>.
- F. Charot and O. Sentieys are members of the steering committee of a CNRS Spring School for graduate students on embedded systems architectures and associated design tools (ARCHI).
- C. Killian was Co-Organizer of the Thematic Day on "Emerging Interconnect Technologies in Many Core Architectures" of GDR SoC<sup>2</sup>, November 27, 2017.

- O. Sentieys is a member of the steering committee of a CNRS spring school for graduate students on low-power design (ECOFAC).
- O. Sentieys is a member of the steering committee of GDR SoC<sup>2</sup>.

#### 9.1.6. Scientific Expertise

- E. Casseau served as an expert for the Natural Sciences and Engineering Research Council of Canada (NSERC), program Discovery Grant 2017.
- O. Sentieys served as a jury member in the EDAA Outstanding Dissertations Award (ODA).

#### 9.2. Teaching - Supervision - Juries

#### 9.2.1. Teaching

- E. Casseau: signal processing, 16h, ENSSAT (L3)
- E. Casseau: low power design, 6h, ENSSAT (M1)
- E. Casseau: real time design methodology, 24h, ENSSAT (M1)
- E. Casseau: computer architecture, 24h, ENSSAT (M1)
- E. Casseau: SoC and high-level synthesis, 24h, Master by Research (SISEA) and ENSSAT (M2)
- S. Derrien: component and system synthesis, 20h, Master by Research (ISTIC) (M2)
- S. Derrien: computer architecture, 12h, ENS Rennes (L3)
- S. Derrien: computer architecture, 24h, ISTIC (L3)
- S. Derrien: introduction to operating systems, 8h, ISTIC (M1)
- S. Derrien: embedded architectures, 48h, ISTIC (M1)
- S. Derrien: high-level synthesis, 6h, ISTIC (M1)
- S. Derrien: software engineering project, 40h, ISTIC (M1)
- F. Charot: processor architecture, 25h, Univ. of Science and Tech. of Hanoi (M1)
- D. Chillet: embedded processor architecture, 20h, ENSSAT (M1)
- D. Chillet: multimedia processor architectures, 24h, ENSSAT (M2)
- D. Chillet: low-power digital CMOS circuits, 6h, Telecom Bretagne (M2)
- C. Killian: digital electronics, 62h, IUT Lannion (L1)
- C. Killian: signal processing, 36h, IUT Lannion (L2)
- C. Killian: automated measurements, 56h, IUT Lannion (L2)
- C. Killian: measurement chain, 58h, IUT Lannion (L2)
- C. Killian: embedded systems programming, 12h, IUT Lannion (L2)
- C. Killian: automatic control, 18h, IUT Lannion (L2)
- A. Kritikakou: computer architecture 1, 32h, ISTIC (L3)
- A. Kritikakou: computer architecture 2, 44h, ISTIC (L3)
- A. Kritikakou: C and unix programming languages, 102h, ISTIC (L3)
- A. Kritikakou: operating systems, 96h, ISTIC (L3)
- A. Kritikakou: multitasking operating systems, 20h, ISTIC (M1)
- O. Sentieys: digital signal processing, 40h, ENSSAT (M1)
- O. Sentieys: VLSI integrated circuit design, 40h, ENSSAT (M1)
- C. Wolinski: computer architectures, 92h, ESIR (L3)
- C. Wolinski: design of embedded systems, 48h, ESIR (M1)
- C. Wolinski: signal, image, architecture, 26h, ESIR (M1)

C. Wolinski: programmable architectures, 10h, ESIR (M1)

C. Wolinski: component and system synthesis, 10h, Master by Research (ISTIC) (M2)

#### 9.2.2. Teaching Responsibilities

- C. Wolinski is the Director of ESIR.
- S. Derrien was the responsible of the first year (M1) of the Master of Computer Science at ISTIC until Aug. 2017.
- O. Sentieys is responsible of the "Embedded Systems" major of the SISEA Master by Research.
- D. Chillet is the responsible of the ICT Master of University of Science and Technology of Hanoi.
- C. Killian is the responsible of the second year of the Physical Measurement DUT at IUT of Lannion.

ENSSAT stands for "École Nationale Supérieure des Sciences Appliquées et de Technologie" and is an "École d'Ingénieurs" of the University of Rennes 1, located in Lannion.

ISTIC is the Electrical Engineering and Computer Science Department of the University of Rennes 1. ESIR stands for "*École supérieure d'ingénieur de Rennes*" and is an "*École d'Ingénieurs*" of the University of Rennes 1, located in Rennes.

#### 9.2.3. Supervision

PhD: Benjamin Barrois, Methods to Evaluate Accuracy-Energy Trade-Off in Operator-Level Approximate Computing, Dec. 2017, O. Sentieys.

PhD: Gaël Deest, Implementation Trade-Offs for FPGA Accelerators, Dec. 2017, S. Derrien.

PhD: Xuan Chien Le, Improving performance of non-intrusive load monitoring with low-cost sensor networks, Apr. 2017, O. Sentieys, B. Vrigneau.

PhD: Rengarajan Ragavan, Error handling and energy estimation for error resilient near-threshold computing, Sep. 2017, O. Sentieys, C. Killian.

PhD: Baptiste Roux, Methodology and Tools for Energy-aware Task Mapping on Heterogeneous Multiprocessor Architectures, Nov. 2017, O. Sentieys, M. Gautier.

PhD in progress: Minh Thanh Cong, Hardware Accelerated Simulation of Heterogeneous Multicore Platforms, May 2017, F. Charot, S. Derrien.

PhD in progress: Petr Dobias, Towards efficient application execution on resilient multi-core architectures, Oct. 2017, E. Casseau.

PhD in progress: Gabriel Gallin, Hardware Arithmetic Units and Crypto-Processor for Hyperelliptic Curves Cryptography, Oct. 2014, A. Tisserand.

PhD in progress: Aymen Gammoudi, New Visual Adaptive Real-Time OS for Embedded Multi-Core Architecture, Oct. 2015, D. Chillet, M.Khalgui.

PhD in progress: Mael Gueguen, Improving the performance and energy efficiency of complex heterogeneous manycore architectures with on-chip data mining, Nov. 2016, O. Sentieys, A. Termier.

PhD in progress: Van-Phu Ha, Application-Level Tuning of Accuracy, Nov. 2017, T. Yuki, O. Sentieys.

PhD in progress: Audrey Lucas, Software support resistant to passive and active attacks for asymmetric cryptography on (very) small computation cores, Jan. 2016, A. Tisserand.

PhD in progress: Jiating, Luo, Communication protocol exploration in the context of 3D integration of multiprocessors interconnected by Optical Network-on-Chip with energy constraints, Nov. 2014, D. Chillet, C. Killian, S. Le-Beux.

PhD in progress: Thibaut Marty, Compiler support for speculative custom hardware accelerators, Sep. 2017, T. Yuki, O. Sentieys.

PhD in progress: Genevieve Ndour, Approximate Computing with High Energy Efficiency for Internet of Things Applications, Apr. 2016, A. Tisserand, A. Molnos (CEA LETI).

PhD in progress: Joel Ortiz Sosa, Study and design of a digital baseband transceiver for wireless network-on-chip architectures, Nov. 2016, O. Sentieys, C. Roland (Lab-STICC).

PhD in progress: Van Dung Pham, Design space exploration in the context of 3D integration of multiprocessors interconnected by Optical Network-on-Chip, Dec 2014, O. Sentieys, D. Chillet, C. Killian, S. Le-Beux.

PhD in progress: Rafail Psiakis, A Self-Healing Reconfigurable Accelerator Structure for Fault-Tolerant Multi-Cores, Oct. 2015, A. Kritikakou, O. Sentieys.

PhD in progress: Simon Rokicki, Hybrid Hardware/Software Dynamic Compilation for Adaptive Embedded Systems, Oct. 2015, S. Derrien.

PhD in progress: Nicolas Roux, Sensor-aided Non-Intrusive Appliance Load Monitoring: Detecting Activity of Devices through Low-Cost Wireless Sensors, Oct. 2016, O. Sentieys, B. Vrigneau.

PhD in progress: Mai-Thanh Tran, Hardware Synthesis of Flexible and Reconfigurable Radio from High-Level Language Dedicated to Physical Layer of Wireless Systems, Oct. 2013, E. Casseau, M. Gautier.

# **10. Bibliography**

#### Major publications by the team in recent years

- [1] R. DAVID, S. PILLEMENT, O. SENTIEYS. *Energy-Efficient Reconfigurable Processors*, in "Low Power Electronics Design", C. PIGUET (editor), Computer Engineering, Vol 1, CRC Press, August 2004, chap. 20.
- [2] S. DERRIEN, S. RAJOPADHYE, P. QUINTON, T. RISSET.12, in "High-Level Synthesis From Algorithm to Digital Circuit", P. COUSSY, A. MORAWIEC (editors), Springer Netherlands, 2008, p. 215-230, http://dx.doi. org/10.1007/978-1-4020-8588-8.
- [3] J.-M. JÉZÉQUEL, B. COMBEMALE, S. DERRIEN, C. GUY, S. RAJOPADHYE.*Bridging the Chasm Between MDE and the World of Compilation*, in "Journal of Software and Systems Modeling (SoSyM)", October 2012, vol. 11, n<sup>0</sup> 4, p. 581-597 [DOI: 10.1007/s10270-012-0266-8], https://hal.inria.fr/hal-00717219.
- [4] B. LE GAL, E. CASSEAU, S. HUET.Dynamic Memory Access Management for High-Performance DSP Applications Using High-Level Synthesis, in "IEEE Transactions on VLSI Systems", 2008, vol. 16, n<sup>o</sup> 11, p. 1454-1464.
- [5] K. MARTIN, C. WOLINSKI, K. KUCHCINSKI, A. FLOCH, F. CHAROT. Constraint Programming Approach to Reconfigurable Processor Extension Generation and Application Compilation, in "ACM transactions on Reconfigurable Technology and Systems (TRETS)", June 2012, vol. 5, n<sup>o</sup> 2, p. 1-38, http://doi.acm.org/10. 1145/2209285.2209289.
- [6] D. MENARD, D. CHILLET, F. CHAROT, O. SENTIEYS. Automatic Floating-point to Fixed-point Conversion for DSP Code Generation, in "Proc. ACM/IEEE CASES", October 2002.
- [7] S. PILLEMENT, O. SENTIEYS, R. DAVID. DART: A Functional-Level Reconfigurable Architecture for High Energy Efficiency, in "EURASIP Journal on Embedded Systems (JES)", 2008, p. 1-13.

- [8] R. ROCHER, D. MENARD, O. SENTIEYS, P. SCALART. Analytical Approach for Numerical Accuracy Estimation of Fixed-Point Systems Based on Smooth Operations, in "IEEE Transactions on Circuits and Systems. Part I, Regular Papers", October 2012, vol. 59, n<sup>o</sup> 10, p. 2326 - 2339 [DOI: 10.1109/TCSI.2012.2188938], http://hal.inria.fr/hal-00741741.
- [9] C. WOLINSKI, M. GOKHALE, K. MCCABE. *A polymorphous computing fabric*, in "IEEE Micro", 2002, vol. 22, n<sup>o</sup> 5, p. 56–68.
- [10] C. WOLINSKI, K. KUCHCINSKI, E. RAFFIN. Automatic Design of Application-Specific Reconfigurable Processor Extensions with UPaK Synthesis Kernel, in "ACM Trans. on Design Automation of Elect. Syst.", 2009, vol. 15, n<sup>o</sup> 1, p. 1–36, http://doi.acm.org/10.1145/1640457.1640458.

#### **Publications of the year**

#### **Doctoral Dissertations and Habilitation Theses**

- [11] B. BARROIS.*Methods to Evaluate Accuracy-Energy Trade-Off in Operator-Level Approximate Computing*, Université de Rennes 1, December 2017, https://hal.inria.fr/tel-01665015.
- [12] G. DEEST. Implementation Trade-Offs for FGPA accelerators, Université de Rennes 1 [UR1], December 2017, https://tel.archives-ouvertes.fr/tel-01665020.
- [13] X.-C. LE. Improving performance of non-intrusive load monitoring with low-cost sensor networks, Université Rennes 1, April 2017, https://tel.archives-ouvertes.fr/tel-01622355.
- [14] R. RAGAVAN. Error handling and energy estimation for error resilient near-threshold computing, Université Rennes 1, September 2017, https://tel.archives-ouvertes.fr/tel-01654476.
- [15] R. RAGAVAN. Error Handling and Energy Estimation Framework For Error Resilient Near-Threshold Computing, Rennes 1, September 2017, https://hal.inria.fr/tel-01636803.
- [16] B. ROUX. Methodology and Tools for Energy-aware Task Mapping on Heterogeneous Multiprocessor Architectures, Université de Rennes 1, November 2017, https://hal.inria.fr/tel-01672814.

#### **Articles in International Peer-Reviewed Journal**

- [17] A. DORAI, V. FRESSE, C. COMBES, E.-B. BOURENNANE, A. MTIBAA. *A collision management structure for NoC deployment on multi-FPGA*, in "Microprocessors and Microsystems: Embedded Hardware Design (MICPRO)", March 2017, vol. 49, p. 28 43 [DOI : 10.1016/J.MICPRO.2017.01.006], https://hal-univ-bourgogne.archives-ouvertes.fr/hal-01484378.
- [18] M. FYRBIAK, S. ROKICKI, N. BISSANTZ, R. TESSIER, C. PAAR. Hybrid Obfuscation to Protect against Disclosure Attacks on Embedded Microprocessors, in "IEEE Transactions on Computers", 2017, https://hal. inria.fr/hal-01426565.
- [19] A. KRITIKAKOU, T. MARTY, M. ROY.DYNASCORE: DYNAmic Software Controller to increase REsource utilization in mixed-critical systems, in "ACM Transactions on Design Automation of Electronic Systems (TODAES)", September 2017, vol. 23, n<sup>o</sup> 2, art ID n°13 [DOI : 10.1145/3110222], https://hal.archivesouvertes.fr/hal-01559696.

- [20] H. LI, S. LE BEUX, M. J. SEPULVEDA FLOREZ, I. O'CONNOR. Energy-Efficiency Comparison of Multi-Layer Deposited Nanophotonic Crossbar Interconnects, in "ACM Journal on Emerging Technologies in Computing Systems", July 2017, vol. XX, https://hal.inria.fr/hal-01508192.
- [21] T. H. NGUYEN, M. GAY, F. GOMEZ AGIS, S. LOBO, O. SENTIEYS, J.-C. SIMON, C. PEUCHERET, L. BRAMERIE.*Impact of ADC parameters on linear optical sampling systems*, in "Optics Communications", November 2017, vol. 402, p. 362-367 [DOI: 10.1016/J.OPTCOM.2017.06.013], https://hal.archivesouvertes.fr/hal-01576164.
- [22] T. H. NGUYEN, P. SCALART, M. GAY, L. BRAMERIE, O. SENTIEYS, J.-C. SIMON, C. PEUCHERET, M. JOINDOT.Blind transmitter IQ imbalance compensation in M-QAM optical coherent systems, in "Journal of optical communications and networking", September 2017, vol. 9, n<sup>o</sup> 9, p. D42-D50 [DOI: 10.1364/JOCN.9.000D42], https://hal.archives-ouvertes.fr/hal-01573632.
- [23] B. ROUXEL, S. DERRIEN, I. PUAUT. Tightening Contention Delays While Scheduling Parallel Applications on Multi-core Architectures, in "ACM Transactions on Embedded Computing Systems (TECS)", October 2017, vol. 16, n<sup>o</sup> 5s, p. 1 - 20 [DOI: 10.1145/3126496], https://hal.archives-ouvertes.fr/hal-01655383.
- [24] C. XIAO, S. WANG, W. LIU, E. CASSEAU. Parallel Custom Instruction Identification for Extensible Processors, in "Journal of Systems Architecture", May 2017, vol. 76, p. 149-159 [DOI: 10.1016/J.SYSARC.2016.11.011], https://hal.inria.fr/hal-01587020.

#### **International Conferences with Proceedings**

- [25] L. AUDREY, A. TISSERAND. ECC Protections against both Observation and Pertubation Attacks, in "CryptArchi 2017: 15th International Workshops on Cryptographic architectures embedded in logic devices", Smolenice, Slovakia, June 2017, https://hal.archives-ouvertes.fr/hal-01545752.
- [26] B. BARROIS, O. SENTIEYS. Customizing Fixed-Point and Floating-Point Arithmetic A Case Study in K-Means Clustering, in "SiPS 2017 - IEEE International Workshop on Signal Processing Systems", Lorient, France, October 2017, https://hal.inria.fr/hal-01633723.
- [27] B. BARROIS, O. SENTIEYS, D. MENARD. The Hidden Cost of Functional Approximation Against Careful Data Sizing – A Case Study, in "Design, Automation & Test in Europe Conference & Exhibition (DATE 2017)", Lausanne, France, 2017, https://hal.inria.fr/hal-01423147.
- [28] T. BOLLENGIER, L. LAGADEC, M. NAJEM, J.-C. LE LANN, P. GUILLOUX. Soft timing closure for soft programmable logic cores: The ARGen approach, in "ARC 2017 - 13th International Symposium on Applied Reconfigurable Computing", Delft, Netherlands, Delft University of Technology, April 2017, https://hal. archives-ouvertes.fr/hal-01475251.
- [29] S. CHERUBIN, G. AGOSTA, I. LASRI, E. ROHOU, O. SENTIEYS. *Implications of Reduced-Precision Computations in HPC: Performance, Energy and Error*, in "International Conference on Parallel Computing (ParCo)", Bologna, Italy, September 2017, https://hal.inria.fr/hal-01633790.
- [30] G. DEEST, T. YUKI, S. RAJOPADHYE, S. DERRIEN. One size does not fit all: Implementation trade-offs for iterative stencil computations on FPGAs, in "FPL - 27th International Conference on Field Programmable Logic and Applications", Gand, Belgium, IEEE, September 2017 [DOI : 10.23919/FPL.2017.8056781], https://hal.inria.fr/hal-01655590.

- [31] S. DERRIEN, I. PUAUT, P. ALEFRAGIS, M. BEDNARA, H. BUCHER, C. DAVID, Y. DEBRAY, U. DURAK, I. FASSI, C. FERDINAND, D. HARDY, A. KRITIKAKOU, G. RAUWERDA, S. REDER, M. SICKS, T. STRIPF, K. SUNESEN, T. TER BRAAK, N. VOROS, J. <sup>†</sup>. BECKER.WCET-aware parallelization of model-based applications for multi-cores: The ARGO approach, in "Design Automation and Test in Europe (DATE), 2017", Lausanne, Switzerland, March 2017, p. 286 289 [DOI: 10.23919/DATE.2017.7927000], http://hal.upmc. fr/hal-01590418.
- [32] A. DORAI, O. SENTIEYS, H. DUBOIS. Evaluation of NoC on Multi-FPGA Interconnection Using GTX Transceiver, in "24th IEEE International Conference on Electronics, Circuits and Systems (ICECS)", Batumi, Georgia, December 2017, https://hal.inria.fr/hal-01633785.
- [33] A. H. EL MOUSSAWI, S. DERRIEN. Superword Level Parallelism aware Word Length Optimization, in "Design, Automation & Test in Europe Conference & Exhibition (DATE 2017)", Lausanne, Switzerland, D. ATIENZA, G. D. NATALE (editors), IEEE, March 2017, https://hal.inria.fr/hal-01425550.
- [34] G. GALLIN, T. OZLUM CELIK, A. TISSERAND.Architecture level Optimizations for Kummer based HECC on FPGAs, in "IndoCrypt 2017 - 18th International Conference on Cryptology in India", Chennai, India, International Conference in Cryptology in India : Progress in Cryptology – INDOCRYPT 2017, Springer, December 2017, vol. 10698, p. 44-64 [DOI: 10.1007/978-3-319-71667-1\_3], https://hal.archives-ouvertes. fr/hal-01614063.
- [35] G. GALLIN, A. TISSERAND. Hardware Architectures for HECC, in "CryptArchi 2017: 15th International Workshops on Cryptographic architectures embedded in logic devices", Smolenice, Slovakia, June 2017, https://hal.archives-ouvertes.fr/hal-01545625.
- [36] G. GALLIN, A. TISSERAND. Hyper-Threaded Multiplier for HECC, in "Asilomar Conference on Signals, Systems, and Computers", Pacific Grove, CA, United States, IEEE, October 2017, https://hal.archivesouvertes.fr/hal-01620046.
- [37] C. KILLIAN, D. CHILLET, S. LE BEUX, O. SENTIEYS, V. D. PHAM, I. O'CONNOR. Energy and Performance Trade-off in Nanophotonic Interconnects using Coding Techniques, in "DAC 2017 - IEEE/ACM Design Automation Conference DAC", Austin, United States, June 2017, 6, https://hal.inria.fr/hal-01495468.
- [38] R. KUMAR BUDHWANI, R. RAGAVAN, O. SENTIEYS. Taking Advantage of Correlation in Stochastic Computing, in "ISCAS 2017 - IEEE International Symposium on Circuits and Systems", Baltimore, United States, May 2017, https://hal.inria.fr/hal-01633725.
- [39] J. LUO, A. ELANTABLY, D. D. PHAM, C. KILLIAN, D. CHILLET, S. LE BEUX, O. SENTIEYS, I. O'CONNOR.*Performance and Energy Aware Wavelength Allocation on Ring-Based WDM 3D Optical NoC*, in "Design, Automation & Test in Europe Conference & Exhibition (DATE 2017)", Lausanne, Switzerland, March 2017, https://hal.inria.fr/hal-01416958.
- [40] L. MO, A. KRITIKAKOU, O. SENTIEYS. Decomposed Task Mapping to Maximize QoS in Energy-Constrained Real-Time Multicores, in "35th IEEE International Conference on Computer Design (ICCD)", Boston, United States, IEEE, November 2017, 6, https://hal.inria.fr/hal-01633782.
- [41] R. PSIAKIS, A. KRITIKAKOU, O. SENTIEYS. NEDA: NOP Exploitation with Dependency Awareness for Reliable VLIW Processors, in "ISVLSI 2017 - IEEE Computer Society Annual Symposium on VLSI", Bochum, Germany, May 2017, p. 391-396 [DOI: 10.1109/ISVLSI.2017.75], https://hal.inria.fr/hal-01633770.

- [42] R. PSIAKIS, A. KRITIKAKOU, O. SENTIEYS. Run-Time Instruction Replication for Permanent and Soft Error Mitigation in VLIW Processors, in "NEWCAS 2017 - 15th IEEE International New Circuits and Systems Conference", Strasbourg, France, June 2017, p. 321-324 [DOI: 10.1109/NEWCAS.2017.8010170], https://hal. inria.fr/hal-01633778.
- [43] R. RAGAVAN, B. BARROIS, C. KILLIAN, O. SENTIEYS. Pushing the Limits of Voltage Over-Scaling for Error-Resilient Applications, in "Design, Automation & Test in Europe Conference & Exhibition (DATE 2017)", Lausanne, Switzerland, March 2017, https://hal.archives-ouvertes.fr/hal-01417665.
- [44] S. ROKICKI, E. ROHOU, S. DERRIEN. Hardware-Accelerated Dynamic Binary Translation, in "IEEE/ACM Design, Automation & Test in Europe Conference & Exhibition (DATE)", Lausanne, Switzerland, March 2017, https://hal.inria.fr/hal-01423639.
- [45] B. ROUXEL, S. DERRIEN, I. PUAUT. *Tightening contention delays while scheduling parallel applications on multi-core architectures*, in "International Conference on Embedded Software (EMSOFT), 2017", Seoul, South Korea, International Conference on Embedded Software, October 2017, 20 [DOI: 10.1145/3126496], http://hal.upmc.fr/hal-01590508.
- [46] Y. UGUEN, F. DE DINECHIN, S. DERRIEN.Bridging High-Level Synthesis and Application-Specific Arithmetic: The Case Study of Floating-Point Summations, in "27th International Conference on Field-Programmable Logic and Applications (FPL)", Gent, Belgium, IEEE, September 2017, 8, https://hal.inria.fr/ hal-01373954.
- [47] I. WALI, E. CASSEAU, A. TISSERAND. An Efficient Framework for Design and Assessment of Arithmetic Operators with Reduced-Precision Redundancy, in "Conference on Design and Architectures for Signal and Image Processing (DASIP)", Dresden, Germany, September 2017, https://hal.inria.fr/hal-01586983.

#### **National Conferences with Proceeding**

[48] D. P. VAN, D. CHILLET, C. KILLIAN, O. SENTIEYS, S. LE BEUX, I. O'CONNOR. *Electrical to Optical Interface for ONoC*, in "GRETSI 2017 - XXVIème colloque", Juan les Pins, France, September 2017, p. 1-4, https://hal.inria.fr/hal-01655417.

#### **Conferences without Proceedings**

[49] G. GALLIN, A. TISSERAND. Hardware Architectures Exploration for Hyper-Elliptic Curve Cryptography, in "Crypto'Puces 2017- 6ème rencontre Crypto'Puces, du composant au système communicant embarqué", Porquerolles, France, May 2017, 31, https://hal.archives-ouvertes.fr/hal-01547034.

#### **Research Reports**

[50] H. YVIQUEL, A. SANCHEZ, R. MICKAËL, E. CASSEAU. *Multicore Runtime for Dynamic Dataflow Video Decoders*, IETR/INSA Rennes ; IRISA, Inria Rennes, April 2017, https://hal.archives-ouvertes.fr/hal-01503378.

#### **Other Publications**

[51] D. CHILLET, D. P. VAN, C. KILLIAN, O. SENTIEYS, S. LE BEUX, I. O'CONNOR. Integration of an Optical NoC into multicore architecture, June 2017, p. 1-2, 2017 - XIIème Colloque National du GDR SoC-SiP, Poster, https://hal.inria.fr/hal-01655420.

- [52] P. DOBIAS, E. CASSEAU, O. SINNEN. Poster: Fault-Tolerant Multi-Processor Scheduling with Backup Copy Technique, September 2017, Conference on Design and Architectures for Signal and Image Processing (DASIP), Poster, https://hal.inria.fr/hal-01610745.
- [53] G. GALLIN, A. TISSERAND.*Finite Field Multiplier Architectures for Hyper-Elliptic Curve Cryptography*, June 2017, Colloque National du GDR SOC2, Poster, https://hal.archives-ouvertes.fr/hal-01539852.
- [54] S. ROKICKI, E. ROHOU, S. DERRIEN. Supporting Runtime Reconfigurable VLIWs Cores Through Dynamic Binary Translation, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01653110.
- [55] Y. UGUEN, F. DE DINECHIN, S. DERRIEN.A high-level synthesis approach optimizing accumulations in floating-point programs using custom formats and operators, January 2017, working paper or preprint, https:// hal.archives-ouvertes.fr/hal-01498357.
- [56] Y. UGUEN, F. DE DINECHIN, S. DERRIEN. *High-Level Synthesis Using Application-Specific Arithmetic: A Case Study*, April 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01502644.

#### **References in notes**

- [57] S. HAUCK, A. DEHON (editors). *Reconfigurable Computing: The Theory and Practice of FPGA-Based Computation*, Morgan Kaufmann, 2008.
- [58] V. BAUMGARTE, G. EHLERS, F. MAY, A. NÜCKEL, M. VORBACH, M. WEINHARDT. PACT XPP A Self-Reconfigurable Data Processing Architecture, in "The Journal of Supercomputing", 2003, vol. 26, n<sup>o</sup> 2, p. 167–184.
- [59] C. BECKHOFF, D. KOCH, J. TORRESEN. Portable module relocation and bitstream compression for Xilinx FPGAs, in "24th Int. Conf. on Field Programmable Logic and Applications (FPL)", 2014, p. 1–8.
- [60] C. BOBDA. Introduction to Reconfigurable Comp.: Architectures Algorithms and Applications, Springer, 2007.
- [61] S. BORKAR, A. A. CHIEN. The Future of Microprocessors, in "Commun. ACM", May 2011, vol. 54, n<sup>o</sup> 5, p. 67–77, http://doi.acm.org/10.1145/1941487.1941507.
- [62] J. M. P. CARDOSO, P. C. DINIZ, M. WEINHARDT. Compiling for reconfigurable computing: A survey, in "ACM Comput. Surv.", June 2010, vol. 42, 13:1, http://doi.acm.org/10.1145/1749603.1749604.
- [63] K. COMPTON, S. HAUCK.*Reconfigurable computing: a survey of systems and software*, in "ACM Comput. Surv.", 2002, vol. 34, n<sup>o</sup> 2, p. 171–210, http://doi.acm.org/10.1145/508352.508353.
- [64] J. CONG, H. HUANG, C. MA, B. XIAO, P. ZHOU.A Fully Pipelined and Dynamically Composable Architecture of CGRA, in "IEEE Int. Symp. on Field-Program. Custom Comput. Machines (FCCM)", 2014, p. 9–16, http://dx.doi.org/10.1109/FCCM.2014.12.
- [65] G. CONSTANTINIDES, P. CHEUNG, W. LUK. Wordlength optimization for linear digital signal processing, in "IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems", October 2003, vol. 22, n<sup>o</sup> 10, p. 1432-1442.

- [66] M. COORS, H. KEDING, O. LUTHJE, H. MEYR. *Fast Bit-True Simulation*, in "Proc. ACM/IEEE Design Automation Conference (DAC)", Las Vegas, june 2001, p. 708-713.
- [67] R. H. DENNARD, F. H. GAENSSLEN, V. L. RIDEOUT, E. BASSOUS, A. R. LEBLANC. Design of ionimplanted MOSFET's with very small physical dimensions, in "IEEE Journal of Solid-State Circuits", 1974, vol. 9, n<sup>o</sup> 5, p. 256–268.
- [68] A. HORMATI, M. KUDLUR, S. MAHLKE, D. BACON, R. RABBAH. Optimus: efficient realization of streaming applications on FPGAs, in "Proc. ACM/IEEE CASES", 2008, p. 41–50.
- [69] H. KALTE, M. PORRMANN. REPLICA2Pro: Task Relocation by Bitstream Manipulation in Virtex-II/Pro FPGAs, in "3rd Conference on Computing Frontiers (CF)", 2006, p. 403–412.
- [70] J.-E. LEE, K. CHOI, N. D. DUTT. Compilation Approach for Coarse-Grained Reconfigurable Architectures, in "IEEE Design and Test of Computers", 2003, vol. 20, n<sup>o</sup> 1, p. 26-33, http://doi.ieeecomputersociety.org/ 10.1109/MDT.2003.1173050.
- [71] H. LEE, D. NGUYEN, J.-E. LEE. Optimizing Stream Program Performance on CGRA-based Systems, in "52nd IEEE/ACM Design Automation Conference", 2015, p. 110:1–110:6, http://doi.acm.org/10.1145/2744769. 2744884.
- [72] B. MEI, S. VERNALDE, D. VERKEST, H. DE MAN, R. LAUWEREINS. ADRES: An architecture with tightly coupled VLIW processor and coarse-grained reconfigurable matrix, in "Proc. FPL", Springer, 2003, p. 61–70.
- [73] N. R. MINISKAR, S. KOHLI, H. PARK, D. YOO.Retargetable Automatic Generation of Compound Instructions for CGRA Based Reconfigurable Processor Applications, in "Proc. ACM/IEEE CASES", 2014, p. 4:1–4:9, http://doi.acm.org/10.1145/2656106.2656125.
- [74] Y. PARK, H. PARK, S. MAHLKE.CGRA express: accelerating execution using dynamic operation fusion, in "Proc. Int. Conf. on Compilers, Architecture, and Synthesis for Embedded Systems", New York, NY, USA, CASES'09, ACM, 2009, p. 271–280, http://doi.acm.org/10.1145/1629395.1629433.
- [75] A. PUTNAM, A. CAULFIELD, E. CHUNG, D. CHIOU, K. CONSTANTINIDES, J. DEMME, H. ES-MAEILZADEH, J. FOWERS, G. GOPAL, J. GRAY, M. HASELMAN, S. HAUCK, S. HEIL, A. HORMATI, J.-Y. KIM, S. LANKA, J. LARUS, E. PETERSON, S. POPE, A. SMITH, J. THONG, P. XIAO, D. BURGER. A reconfigurable fabric for accelerating large-scale datacenter services, in "ACM/IEEE 41st International Symposium on Computer Architecture (ISCA)", June 2014, p. 13-24, http://dx.doi.org/10.1109/ISCA.2014.6853195.
- [76] G. THEODORIDIS, D. SOUDRIS, S. VASSILIADIS.2, in "A survey of coarse-grain reconfigurable architectures and CAD tools", Springer Verlag, 2007.
- [77] G. VENKATARAMANI, W. NAJJAR, F. KURDAHI, N. BAGHERZADEH, W. BOHM, J. HAMMES. Automatic compilation to a coarse-grained reconfigurable system-on-chip, in "ACM Trans. on Emb. Comp. Syst.", 2003, vol. 2, n<sup>O</sup> 4, p. 560–589, http://doi.acm.org/10.1145/950162.950167.

# **Project-Team CELTIQUE**

# Software certification with semantic analysis

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: Université Rennes 1

École normale supérieure de Rennes

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Proofs and Verification

# **Table of contents**

1.	Personnel	133
2.	Overall Objectives	134
3.	New Software and Platforms	134
	3.1. Jacal	134
	3.2. Javalib	135
	3.3. JSCert	135
	3.4. SAWJA	135
	3.5. Timbuk	135
4.	New Results	136
	4.1. Higher-Order Process Calculi	136
	4.2. Certified Semantics and Analyses for JavaScript	136
	4.3. Certified Concurrent Garbage Collector	136
	4.4. Static analysis of functional programs using tree automata and term rewriting	137
	4.5. C Semantics and Certified Compilation	137
	4.6. Constant-time verification by compilation and static analysis	137
5.	Partnerships and Cooperations	<b>138</b>
	5.1. National Initiatives	138
	5.1.1. The ANR AnaStaSec project	138
	5.1.2. The ANR Binsec project	138
	5.1.3. The ANR MALTHY project	138
	5.1.4. The ANR AJACS project	138
	5.1.5. The ANR DISCOVER project	139
	5.2. European Initiatives	139
	5.3. International Initiatives	140
6.	Dissemination	140
	6.1. Promoting Scientific Activities	140
	6.1.1. Scientific Events Selection	140
	6.1.1.1. Chair of Conference Program Committees	140
	6.1.1.2. Member of the Conference Program Committees	140
	6.1.1.3. Reviewer	141
	6.1.2. Journal	141
	6.1.3. Invited Talks	141
	6.1.4. Scientific Expertise	141
	6.1.5. Research Administration	141
	6.2. Teaching - Supervision - Juries	141
	6.2.1. Teaching	141
	6.2.2. Supervision	142
	6.2.3. Juries	143
	6.3. Popularization	143
7.	Bibliography	144

# **Project-Team CELTIQUE**

Creation of the Project-Team: 2009 July 01

#### **Keywords:**

#### **Computer Science and Digital Science:**

- A2.1. Programming Languages
- A2.1.1. Semantics of programming languages
- A2.1.2. Object-oriented programming
- A2.1.3. Functional programming
- A2.1.9. Dynamic languages
- A2.2. Compilation
- A2.2.1. Static analysis
- A2.2.2. Memory models
- A2.4. Verification, reliability, certification
- A2.4.1. Analysis
- A2.4.2. Model-checking
- A2.4.3. Proofs
- A4. Security and privacy
- A4.5. Formal methods for security

#### **Other Research Topics and Application Domains:**

- B6.1. Software industry
- B6.1.1. Software engineering
- B6.6. Embedded systems

# 1. Personnel

#### **Research Scientists**

Thomas Jensen [Team leader, Inria, Senior Researcher, HDR] Frédéric Besson [Inria, Researcher] Alan Schmitt [Inria, Senior Researcher, HDR]

#### **Faculty Members**

Sandrine Blazy [Univ de Rennes I, Professor, HDR] David Cachera [Ecole normale supérieure de Rennes, Associate Professor, HDR] Delphine Demange [Univ de Rennes I, Associate Professor] Thomas Genet [Univ de Rennes I, Associate Professor, HDR] Serguei Lenglet [Univ de Lorraine, Associate Professor] David Pichardie [Ecole normale supérieure de Rennes, Professor, HDR]

#### **Technical Staff**

Yannick Zakowski [Univ de Rennes I, from Dec 2017]

#### **PhD Students**

Pauline Bolignano [Prove & Run, until Feb 2017] Gurvan Cabon [Inria] Alexandre Dang [Inria] Yon Fernandez de Retana [Univ de Rennes I]

```
Timothée Haudebourg [Univ de Rennes I, from Oct 2017]
Julien Lepiller [Inria]
Florent Saudel [Amossys]
Alix Trieu [Univ de Rennes I]
Yannick Zakowski [Ecole normale supérieure de Rennes, until Nov 2017]
```

Visiting Scientist

Ahmad Salim Al-Sibahi [IT University of Copenhagen, until Jan 2017]

Administrative Assistant Lydie Mabil [Inria]

# 2. Overall Objectives

#### 2.1. Project overview

The overall goal of the CELTIQUE project is to improve the security and reliability of software with semanticsbased modeling, analysis and certification techniques. To achieve this goal, the project conducts work on improving semantic description and analysis techniques, as well as work on using proof assistants (most notably Coq) to develop and prove properties of these techniques. We are applying such techniques to a variety of source languages, including Java, C, and JavaScript. We also study how these techniques apply to low-level languages, and how they can be combined with certified compilation. The CompCert certified compiler and its intermediate representations are used for much of our work on semantic modeling and analysis of C and lower-level representations.

The semantic analyses extract approximate but sound descriptions of software behaviour from which a proof of safety or security can be constructed. The analyses of interest include numerical data flow analysis, control flow analysis for higher-order languages, alias and points-to analysis for heap structure manipulation. In particular, we have designed several analyses for information flow control, aimed at computing attacker knowledge and detecting side channels.

We work with three application domains: Java software for small devices (in particular smart cards and mobile telephones), embedded C programs, and web applications.

CELTIQUE is a joint project with the CNRS, the University of Rennes 1 and ENS Rennes.

# 3. New Software and Platforms

#### 3.1. Jacal

#### JAvaCard AnaLyseur

KEYWORDS: JavaCard - Certification - Static program analysis - AFSCM

FUNCTIONAL DESCRIPTION: Jacal is a JAvaCard AnaLyseur developed on top of the SAWJA platform. This proprietary software verifies automatically that Javacard programs conform with the security guidelines issued by the AFSCM (Association Française du Sans Contact Mobile). Jacal is based on the theory of abstract interpretation and combines several object-oriented and numeric analyses to automatically infer sophisticated invariants about the program behaviour. The result of the analysis is thereafter harvest to check that it is sufficient to ensure the desired security properties.

- Participants: David Pichardie, Delphine Demange, Frédéric Besson and Thomas Jensen
- Contact: Thomas Jensen

# 3.2. Javalib

FUNCTIONAL DESCRIPTION: Javalib is an efficient library to parse Java .class files into OCaml data structures, thus enabling the OCaml programmer to extract information from class files, to manipulate and to generate valid .class files.

- Participants: David Pichardie, Frédéric Besson, Laurent Guillo, Laurent Hubert, Nicolas Barré, Pierre Vittet and Tiphaine Turpin
- Contact: Frédéric Besson
- URL: http://sawja.inria.fr/

# 3.3. JSCert

#### Certified JavaScript

FUNCTIONAL DESCRIPTION: The JSCert project aims to really understand JavaScript. JSCert itself is a mechanised specification of JavaScript, written in the Coq proof assistant, which closely follows the ECMAScript 5 English standard. JSRef is a reference interpreter for JavaScript in OCaml, which has been proved correct with respect to JSCert and tested with the Test 262 test suite.

- Participants: Alan Schmitt and Martin Bodin
- Partner: Imperial College London
- Contact: Alan Schmitt
- URL: http://jscert.org/

# 3.4. SAWJA

Static Analysis Workshop for Java

KEYWORDS: Security - Software - Code review - Smart card

SCIENTIFIC DESCRIPTION: Sawja is a library written in OCaml, relying on Javalib to provide a high level representation of Java bytecode programs. It name comes from Static Analysis Workshop for JAva. Whereas Javalib is dedicated to isolated classes, Sawja handles bytecode programs with their class hierarchy and with control flow algorithms.

Moreover, Sawja provides some stackless intermediate representations of code, called JBir and A3Bir. The transformation algorithm, common to these representations, has been formalized and proved to be semantics-preserving.

See also the web page http://sawja.inria.fr/.

Version: 1.5

#### Programming language: Ocaml

FUNCTIONAL DESCRIPTION: Sawja is a toolbox for developing static analysis of Java code in bytecode format. Sawja provides advanced algorithms for reconstructing high-level programme representations. The SawjaCard tool dedicated to JavaCard is based on the Sawja infrastructure and automatically validates the security guidelines issued by AFSCM (http://www.afscm.org/). SawjaCard can automate the code audit process and automatic verification of functional properties.

- Participants: David Pichardie, Frédéric Besson and Laurent Guillo
- Partners: CNRS ENS Cachan
- Contact: Frédéric Besson
- URL: http://sawja.inria.fr/

#### 3.5. Timbuk

KEYWORDS: Demonstration - Ocaml - Vérification de programmes - Tree Automata

FUNCTIONAL DESCRIPTION: Timbuk is a collection of tools for achieving proofs of reachability over Term Rewriting Systems and for manipulating Tree Automata (bottom-up non-deterministic finite tree automata)

RELEASE FUNCTIONAL DESCRIPTION: This version does no longer include the tree automata library but focuses on reachability analysis and equational approximations.

- Participant: Thomas Genet
- Contact: Thomas Genet
- URL: http://www.irisa.fr/celtique/genet/timbuk/

# 4. New Results

#### 4.1. Higher-Order Process Calculi

Participants: Sergueï Lenglet, Alan Schmitt.

Sergueï Lenglet and Alan Schmitt, in collaboration with researchers at Wrocław university, designed a fully abstract encoding of the  $\lambda$ -calculus into HOcore, a minimal higher-order process calculus. This work has been published at LICS [37]. In parallel, Lenglet and Schmitt have formalized HO $\pi$  in Coq and showed that its bisimilarity is compatible using Howe's method. This work has been accepted for publication at CPP 2018 [30].

#### 4.2. Certified Semantics and Analyses for JavaScript

Participants: Gurvan Cabon, Alan Schmitt.

Alan Schmitt has continued his collaboration with Arthur Charguéraud (Inria Nancy) and Thomas Wood (Imperial College London) to develop JSExplain, an interpreter for JavaScript that is as close as possible to the specification. The tool is publicly available at https://github.com/jscert/jsexplain and is being extended to cover the current version of the standard.

In parallel, Gurvan Cabon and Alan Schmitt have developed a framework to automatically derive an information-flow tracking semantics from a pretty-big-step semantics. This work has been published [34] and is being formalized in Coq.

#### 4.3. Certified Concurrent Garbage Collector

Participants: Yannick Zakowski, David Cachera, Delphine Demange, David Pichardie.

Concurrent garbage collection algorithms are an emblematic challenge in the area of concurrent program verification. We addressed this problem by proposing a mechanized proof methodology based on the popular Rely-Guarantee (RG) proof technique. We designed a specific compiler intermediate representation (IR) with strong type guarantees, dedicated support for abstract concurrent data structures, and high-level iterators on runtime internals (objects, roots, fields, thread identifiers...). In addition, we defined an RG program logic supporting an incremental proof methodology where annotations and invariants can be progressively enriched. We have formalized the IR, the proof system, and proved the soundness of the methodology in the Coq proof assistant. Equipped with this IR, we have proved the correctness of a fully concurrent garbage collector where mutators never have to wait for the collector. This work has been published in [32].

In this work, reasoning simultaneously about the garbage collection algorithm and the concrete implementation of the concurrent data-structures it uses would have entailed an undesired and unnecessary complexity. The above proof is therefore conducted with respect to abstract operations which execute atomically. In practice, however, concurrent data-structures uses fine-grained concurrency, for performance reasons. One must therefore prove an observational refinement between the abstract concurrent data-structures and their finedgrained, "linearisable" implementation. To adress this issue, we introduce a methodology inspired by the work of Vafeiadis, and provide the approach with solid semantic foundations. Assuming that fine-grained implementations are proved correct with respect to an RG specification encompassing linearization conditions, we prove, once and for all, that this entails a semantic refinement of their abstraction. This methodology is instantiated to prove correct the main data-structure used in our garbage collector. This work has been published in [33].

# 4.4. Static analysis of functional programs using tree automata and term rewriting

Participants: Thomas Genet, Thomas Jensen, Timothée Haudebourg.

We develop a specific theory and the related tools for analyzing programs whose semantics is defined using term rewriting systems. The analysis principle is based on regular approximations of infinite sets of terms reachable by rewriting. Regular tree languages are (possibly) infinite languages which can be finitely represented using tree automata. To over-approximate sets of reachable terms, the tools we develop use the Tree Automata Completion (TAC) algorithm to compute a tree automaton recognizing a superset of all reachable terms. This over-approximation is then used to prove properties on the program by showing that some "bad" terms, encoding dangerous or problematic configurations, are not in the superset and thus not reachable. This is a specific form of, so-called, Regular Tree Model Checking. We have already shown that tree automata completion can safely over-approximate the image of any first-order complete and terminating functional program. We have extended this result to the case of higher-order functional programs [40] and obtained very encouraging experimental results http://people.irisa.fr/Thomas.Genet/timbuk/funExperiments/. Besides, we have shown that completion was abble to take the evaluation strategy of the program into account [19]. The next step is to show the completeness of the approach, i.e., that any regular approximation of the image of a function can be found using completion. We already made progress in this direction [39].

#### 4.5. C Semantics and Certified Compilation

Participants: Frédéric Besson, Sandrine Blazy.

The COMPCERT C compiler provides the formal guarantee that the observable behaviour of the compiled code improves on the observable behaviour of the source code. A first limitation of this guarantee is that if the source code goes wrong, i.e. does not have a well-defined behaviour, any compiled code is compliant. Another limitation is that COMPCERT 's notion of observable behaviour is restricted to IO events.

Over the past years, we have refined the semantics underlying COMPCERT so that (unlike COMPCERT but like GCC) the binary representation of pointers can be manipulated much like integers and such that memory is a finite resource. We have now a formally verified C compiler, COMPCERTS, which is essentially the COMPCERT compiler, albeit with a stronger formal guarantee. The semantics preservation theorem applies to a wider class of existing C programs and, therefore, their compiled version benefits from the formal guarantee of COMPCERTS. COMPCERTS preserves not only the observable behaviour of programs but also ensures that the memory consumption is preserved by the compiler. As a result, we have the formal guarantee that the compiled code requires no more memory than the source code. This ensures that the absence of stack-overflows is preserved by compilation.

The whole proof of COMPCERTS represents a significant proof-effort. Details about the formal definition of the semantics and the proof of compiler passes can be found in the following publications [17], [25]

#### 4.6. Constant-time verification by compilation and static analysis

Participants: Sandrine Blazy, David Pichardie, Alix Trieu.

To protect their implementations, cryptographers follow a very strict programming discipline called constanttime programming. They avoid branchings controlled by secret data as an attacker could use timing attacks, which are a broad class of side-channel attacks that measure different execution times of a program in order to infer some of its secret values. Several real-world secure C libraries such as NaCl, mbedTLS, or Open Quantum Safe, follow this discipline. We propose an advanced static analysis, based on state-of-the-art techniques from abstract interpretation, to report time leakage during programming. To that purpose, we analyze source C programs and use full context-sensitive and arithmetic-aware alias analyses to track the tainted flows. We give semantic evidences of the correctness of our approach on a core language. We also present a prototype implementation for C programs that is based on the CompCert compiler toolchain and its companion Verasco static analyzer. We present verification results on various real-world constant-time programs and report on a successful verification of a challenging SHA-256 implementation that was out of scope of previous toolassisted approaches. This work has been published at ESORICS'17 [27]. The previous technique is well-adapted to verify the constant-time discipline at source level and give feedback to programmers, but the final security property must be established on the executable form of the program. In a joint work with IMDEA Software (Gilles Barthe and Vincent Laporte), we propose an automated methodology for validating on low-level intermediate representations the results of a source-level static analysis. Our methodology relies on two main ingredients: a relative-safety checker, an instance of a relational verifier which proves that a program is *safer* than another, and a transformation of programs into defensive form which verifies the analysis results at runtime. We prove the soundness of the methodology, and provide a formally verified instantiation based on the Verasco verified C static analyzer and the CompCert verified C compiler. This work has been published at CSF'17 [24].

# 5. Partnerships and Cooperations

#### 5.1. National Initiatives

#### 5.1.1. The ANR AnaStaSec project

Participants: Frédéric Besson, Sandrine Blazy, Thomas Jensen, Alexandre Dang, Julien Lepiller.

Static program analysis, Security, Secure compilation

The AnaStaSec project (2015–2018) aims at ensuring security properties of embedded critical systems using static analysis and security enhancing compiler techniques. The case studies are airborne embedded software with ground communication capabilities. The Celtique project focuses on software fault isolation which is a compiler technology to ensure by construction a strong segregation of tasks.

This is a joint project with the Inria teams ANTIQUE and PROSECCO, CEA-LIST, TrustInSoft, AMOSSYS and Airbus Group.

#### 5.1.2. The ANR Binsec project

Participants: Frédéric Besson, Sandrine Blazy, Pierre Wilke, Julien Lepiller.

Binary code, Static program analysis

The Binsec project (2013–2017) is funded by the call ISN 2012, a program of the Agence Nationale de la Recherche. The goal of the BINSEC project is to develop static analysis techniques and tools for performing automatic security analyses of binary code. We target two main applicative domains: vulnerability analysis and virus detection.

Binsec is a joint project with the Inria CARTE team, CEA LIS, VERIMAG and EADS IW.

#### 5.1.3. The ANR MALTHY project

Participant: David Cachera.

The MALTHY project, funded by ANR in the program INS 2013, aims at advancing the state-of-the-art in realtime and hybrid model checking by applying advanced methods and tools from linear algebra and algebraic geometry. MALTHY is coordinated by VERIMAG, involving CEA-LIST, Inria Rennes (Tamis and Celtique), Inria Saclay (MAXPLUS) and VISEO/Object Direct.

#### 5.1.4. The ANR AJACS project

Participants: Martin Bodin, Gurvan Cabon, Thomas Jensen, Alan Schmitt.

The goal of the AJACS project is to provide strong security and privacy guarantees on the client side for web application scripts. To this end, we propose to define a mechanized semantics of the full JavaScript language, the most widely used language for the Web. We then propose to develop and prove correct analyses for JavaScript programs, in particular information flow analyses that guarantee no secret information is leaked to malicious parties. The definition of sub-languages of JavaScript, with certified compilation techniques targeting them, will allow us to derive more precise analyses. Finally, we propose to design and certify security and privacy enforcement mechanisms for web applications, including the APIs used to program real-world applications.

The project partners include the following Inria teams: Celtique, Indes, Prosecco, and Toccata; it also involves researchers from Imperial College as external collaborators. The project runs from December 2014 to November 2018.

#### 5.1.5. The ANR DISCOVER project

**Participants:** Sandrine Blazy, Delphine Demange, Thomas Jensen, David Pichardie, Yon Fernandez de Retana, Yannick Zakovski.

The DISCOVER project project aims at leveraging recent foundational work on formal verification and proof assistants to design, implement and verify compilation techniques used for high-level concurrent and managed programming languages. The ultimate goal of DISCOVER is to devise new formalisms and proof techniques able to scale to the mechanized correctness proof of a compiler involving a rich class of optimizations, leading to efficient and scalable applications, written in higher-level languages than those currently handled by cutting-edge verified compilers.

In the light of recent work in optimizations techniques used in production compilers of high-level languages, control-flow-graph based intermediate representations seems too rigid. Indeed, the analyses and optimizations in these compilers work on more abstract representations, where programs are represented with data and control dependencies. The most representative representation is the sea-of-nodes form, used in the Java Hotspot Server Compiler, and which is the rationale behind the highly relaxed definition of the Java memory model. DISCOVER proposes to tackle the problem of verified compilation for shared-memory concurrency with a resolute language-based approach, and to investigate the formalization of adequate program intermediate representations and associated correctness proof techniques.

The project runs from October 2014 to September 2019.

#### 5.2. European Initiatives

#### 5.2.1. Collaborations in European Programs, Except FP7 & H2020

Program:CA COST Action CA15123

Project acronym: EUTYPES

Project title: European research network on types for programming and verification

Duration: 03/2016 to 03/2020

Coordinator: Herman Geuvers (Radboud University Nijmegen, The Netherlands)

Other partners: Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Macedonia, Germany, Hungary, Israel, Italy, Lithuania, Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovenia, Spain, Sweden, United Kingdom

Abstract: Types are pervasive in programming and information technology. A type defines a formal interface between software components, allowing the automatic verification of their connections, and greatly enhancing the robustness and reliability of computations and communications. In rich dependent type theories, the full functional specification of a program can be expressed as a type. Type systems have rapidly evolved over the past years, becoming more sophisticated, capturing new aspects of the behaviour of programs and the dynamics of their execution.

This COST Action will give a strong impetus to research on type theory and its many applications in computer science, by promoting (1) the synergy between theoretical computer scientists, logicians and mathematicians to develop new foundations for type theory, for example as based on the recent development of "homotopy type theory", (2) the joint development of type theoretic tools as proof assistants and integrated programming environments, (3) the study of dependent types for programming and its deployment in software development, (4) the study of dependent types for verification and its deployment in software analysis and verification. The action will also tie together these different areas and promote cross-fertilisation.

Sandrine Blazy is Substitute Member of the Managment Committee for France.

#### 5.3. International Initiatives

#### 5.3.1. Inria International Partners

5.3.1.1. Declared Inria International Partners

#### WEBCERT

Title: Verified Trustworthy web Applications

International Partner (Institution - Laboratory - Researcher):

Imperial College London - Department of Computing - Philippa Gardner

Duration: 2015 - 2019

Start year: 2015

See also: JSCert web page

The WebCert partnership focuses on applying formal methods to the JavaScript language: mechanized specification, development of an executable formal specification, design of a program logic, development of verification tools, and study of secure sub-languages.

# 6. Dissemination

#### 6.1. Promoting Scientific Activities

#### 6.1.1. Scientific Events Selection

#### 6.1.1.1. Chair of Conference Program Committees

 CoqPL 2017 (International Workshop on Coq for PL) was chaired by Sandrine Blazy and Emilio Jesus Gallego Arias

#### 6.1.1.2. Member of the Conference Program Committees

- TASE 2017 (Symposium on Theoretical Aspects of Software Engineering): Alan Schmitt
- Web Programming 2018: Alan Schmitt
- ProWeb 2018: Alan Schmitt
- CC 2017 (Conference on Compiler Construction) : David Pichardie
- ESORICS 2017 (European Symposium on Research in Computer Security) : David Pichardie
- ESOP 2017 (European Symposium on Programming) : David Pichardie
- CC 2018 (Conference on Compiler Construction) : David Pichardie
- CoqPL 2017 (International Workshop on Coq for PL) : Sandrine Blazy
- AFADL 2017 (Approches Formelles dans l'Assistance au Développement de Logiciels) : Sandrine Blazy
- SRC (Student Research Competition) @ PLDI 2017 : Sandrine Blazy
- VSTTE 2017 (Verified Software: Theories, Tools, and Experiments) : Sandrine Blazy
- GPCE 2017 (Generative Programming: Concepts & Experiences) : Sandrine Blazy
- IFL 2017 (International symposium on Implementation and application of Functional Languages) : Sandrine Blazy
- TFP 2017 (Trends in Functional Programming) : Sandrine Blazy
- CPP 2018 (ACM SIGPLAN Conference on Certified Programs and Proofs) : Sandrine Blazy
- Euro S&P 2018 (IEEE European Symposium on Security and Privacy) : Sandrine Blazy
- TACAS 2017 (Tools and Algorithms for the Construction and Analysis of Software : Thomas Jensen.

- FCS 2017 (Int. workshop on Foundations of Computer Security) : Thomas Jensen.
- SAS 2017 (Static Analysis Symposium) : Thomas Jensen.

#### 6.1.1.3. Reviewer

• POPL 2018 (Symposium on Principles of Programming Languages): Alan Schmitt

#### 6.1.2. Journal

6.1.2.1. Reviewer - Reviewing Activities

- Information & Computation: Alan Schmitt
- Science of Computer Programming: Alan Schmitt
- Discrete Mathematics & Theoretical Computer: Alan Schmitt
- Theoretical Computer Science: Alan Schmitt
- Journal of Logical and Algebraic Methods in Programming: Alan Schmitt
- ACM Transactions on Privacy and Security (TOPS): David Pichardie

#### 6.1.3. Invited Talks

- Delphine Demange: "On-the-Fly Garbage Collection: An Exercise in Compiler Verification". Inria Scientific Days 2017. June 2017.
- Thomas Genet: "SPAN+AVISPA for Verifying Cryptographic Protocols". RESSI (Rendez-vous de la recherche et de l'enseignement de la sécurité des systèmes d'information), Grenoble, May 2017 [42].
- Thomas Genet: "Tree Automata for Reachability in Rewriting". International School on Rewriting, Eindhoven, July 2017. http://www.win.tue.nl/~hzantema/isr.html.
- Thomas Jensen: Formal methods for software security, Forum Méthodes Formelles, Toulouse, France, Jan. 2017 [21].
- Thomas Jensen: Formal methods for software security, Journée inaugurale GDR Sécurité Informatique, Paris, June 2017 [22].
- Thomas Jensen. Hybrid information flow analysis against web tracking.. The 12th International Conference on Risks and Security of Internet and Systems (CRiSIS 2017), Dinard, France, Sept. 2017 [23].

#### 6.1.4. Scientific Expertise

- Sandrine Blazy: expertise of an ERC Advanced Grant research proposal.
- Thomas Jensen is Inria representative in the European Cyber Security Organisation (ECSO) working group in Research and Innovation.

#### 6.1.5. Research Administration

- Sandrine Blazy is member of Section 6 of the national committee for scientific research CoNRS.
- Sandrine Blazy is coordinator of the LTP (Languages, Types, Proofs) group of the French GDR GPL.
- Thomas Jensen is head of the NUMERIC department at Université Bretagne Loire.
- Thomas Jensen is director of the IT Security track and member of the executive board of the Laboratoire d'excellence "CominLabs".

# 6.2. Teaching - Supervision - Juries

#### 6.2.1. Teaching

Licence : Alan Schmitt, Programmation Fonctionnelle, 36h, L3, Insa Rennes, France

Licence : Delphine Demange, Spécialité Informatique 1 - Algorithmique et Complexité Expérimentale, 36h, L1, Université Rennes 1, France

Licence : Delphine Demange, Spécialité Informatique 2 - Functional and Immutable Programming, 70h, L1, Université Rennes 1, France

Licence : Delphine Demange, Programmation de Confiance, 36h, L3, Université Rennes 1, France Licence : David Pichardie, Graph algorithms, 24h, L3, ESIR, France

Licence : Sandrine Blazy, Functional programming, 30h, L3, Université Rennes 1, France

Licence: Thomas Genet, Software Engineering, 58h, L2, Université de Rennes 1, France

Licence : Thomas Genet, Spécialité Informatique 1 - Algorithmic and Experimental Complexity, 42h, L1, Université Rennes 1, France

Master : Sandrine Blazy, Méthodes Formelles pour le développement de logiciels sûrs, 53h, M1, Université Rennes 1, France

Master : Alan Schmitt, Méthodes Formelles pour le développement de logiciels sûrs, 25h, M1, Université Rennes 1, France France

Master : Sandrine Blazy, Mechanized Semantics, 15h, M2, Université Rennes 1, France

Master : Sandrine Blazy, Semantics, 24h, M1, Université Rennes 1, France

Master : Sandrine Blazy, Software vulnerabilities, 20h, M2, Université Rennes 1, France

Master : Delphine Demange, Software Security, 9h, M2, Université Rennes 1, France

Master : David Cachera, Semantics, 24h, M1, Université Rennes 1, France

Master : David Cachera, Advanced Semantics, 20h, M2, Université Rennes 1, France

Master : Thomas Genet, Formal Design and Verification, 108h, M1, Université de Rennes 1, France.

Master : Thomas Jensen, Program Analysis and Software Security, 21h, M2, Université Rennes 1, France

#### 6.2.2. Supervision

PhD in progress : Timothée Haudebourg, Lightweight Formal Verification for Functional Programs, 1st october 2017, Thomas Genet and Thomas Jensen

PhD in progress : Alexandre Dang, Security by compilation, 1st september 2016, Frédéric Besson and Thomas Jensen

PhD in progress : Julien Lepiller, Binary analysis for Isolation, 1st september 2016, Frédéric Besson and Thomas Jensen

PhD in progress : Gurvan Cabon, Analyse non locale certifiée en JavaScript grâce à une sémantique annotée, 1st september 2015, Alan Schmitt

PhD in progress : Florent Saudel, Vulnerability discovery, November 2015, Sandrine Blazy, Frédéric Besson and Cédric Berthion (Amossys)

PhD in progress : Alix Trieu, Formally verified compilation and static analysis, January 2016, Sandrine Blazy and David Pichardie

PhD in progress : Yon Fernandez De Retana, Verified Optimising Compiler for high-level languages, 1st september 2015, Delphine Demange and David Pichardie

David Bühler, Structuring an abstract interpreter through value and state abstractions, defended March 2017, Sandrine Blazy and Boris Yakobowski (CEA)

Yannick Zakowski, Verification of a Concurrent Garbage Collector, defended December 2017, David Pichardie and David Cachera.

Pauline Bolignano, Formal models and verification of memory management in a hypervisor, defended May 2017, Thomas Jensen and Vincent Siles (Prove & Run). Oana Andreescu, Static analysis of functional programs with an application to the frame problem in deductive verification, May 2017, Thomas Jensen and Stéphane Lescuyer (Prove & Run).

#### 6.2.3. Juries

- Alan Schmitt, jury member for the selection of Inria CR (researcher) candidates, March and April 2017, Inria, Rennes, France.
- Thomas Jensen, jury member for the selection of Inria CR (researcher) candidates, March and April 2017, Inria, Rennes, France.
- Sandrine Blazy, jury member for the selection of CNRS CR and DR (researchers) candidates, February and March 2017, CNRS, Paris, France.
- Sandrine Blazy, jury member for the selection of a professor at University of Copenhagen, May 2017, Copenhagen, Denmark.
- Sandrine Blazy, jury member (reviewer) for the PhD defense of Romain Aïssat, January 2017, Paris-Sud University
- Sandrine Blazy, jury member for the PhD defense of Oana Andreescu, May 2017, Université Rennes 1
- Sandrine Blazy, jury member for the PhD defense of Ninon Eyrolles, June 2017, Université Versailles Saint-Quentin
- Sandrine Blazy, jury member (reviewer) for the HDR defense of Alain Giorgetti, December 2017, Université de Franche-Comté
- Sandrine Blazy, jury member for the PhD defense of Jordy Ruiz, December 2017, Université de Toulouse
- Sandrine Blazy, jury member for the PhD defense of Pierre Lestringant, December 2017, Université Rennes 1.
- Sandrine Blazy, jury member of the GDR GPL PhD award committee.
- David Pichardie, external reviewer for the PhD defense of Hendra Gunadi, July 2017, Australian National University, Canberra, Australia.
- David Pichardie, Licenciate discussion leader for the PhD student Marco Vassena, Chalmers University of Technology, Gothenburg, Sweden.
- Delphine Demange, jury member of the Gilles Kahn PhD award committee, December 2017, Inria Paris
- Delphine Demange, jury member for the PhD defense of Pauline Bolignano, May 2017, Université Rennes 1
- Thomas Genet, jury member (reviewer) for the PhD defense of Vivien Pelletier, October 2017, Université d'Orléans, France.
- Thomas Jensen, jury member for the HdR defense of Charlotte Truchet, November 2017, Université de Nantes, France.
- Thomas Jensen, jury member (reviewer) for the PhD defense of Zeineb Zhioua, September 2017, Téleécom ParisTech, France.
- Thomas Jensen, jury member for the PhD defense of Deepak Subramanian, December 2017, CentraleSupélec, France.

#### 6.3. Popularization

Article "JavaScript, un langage à la croissance organique", Alan Schmitt, blog Binaire Le Monde. http://binaire.blog.lemonde.fr/2017/05/12/javascript-un-langage-a-la-croissance-organique/

Article "L'assistant de preuve Coq", Sandrine Blazy, Pierre Castéran, Hugo Herbelin, Techniques et Sciences de l'ingénieur, août 2017. https://www.techniques-ingenieur.fr/base-documentaire/technologies-de-l-information-th9/programmation-42304210/coq-assistant-de-preuve-h3310/

Talk "Bug, Virus, Intrusion, Pirates... So many threats and no defense? Yes... maths.", Thomas Genet, given three times in high schools close to Rennes.

# 7. Bibliography

#### Major publications by the team in recent years

- [1] G. BARTHE, D. DEMANGE, D. PICHARDIE. Formal Verification of an SSA-based Middle-end for CompCert, in "ACM Transactions on Programming Languages and Systems (TOPLAS)", 2014, 35, https://hal.inria.fr/hal-01097677.
- [2] F. BESSON, N. BIELOVA, T. JENSEN. Hybrid Information Flow Monitoring Against Web Tracking, in "CSF - 2013 IEEE 26th Computer Security Foundations Symposium", New Orleans, United States, 2013 [DOI: 10.1109/CSF.2013.23], http://hal.inria.fr/hal-00924138.
- [3] F. BESSON, T. JENSEN, D. PICHARDIE. Proof-Carrying Code from Certified Abstract Interpretation to Fixpoint Compression, in "Theoretical Computer Science", 2006, vol. 364, n<sup>o</sup> 3, p. 273–291.
- [4] M. BODIN, A. CHARGUÉRAUD, D. FILARETTI, P. GARDNER, S. MAFFEIS, D. NAUDZIUNIENE, A. SCHMITT, G. SMITH. A Trusted Mechanised JavaScript Specification, in "POPL 2014 - 41st ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages", San Diego, United States, November 2013, http://hal.inria.fr/hal-00910135.
- [5] B. BOYER, T. GENET, T. JENSEN. Certifying a Tree Automata Completion Checker, in "4th International Joint Conference, IJCAR 2008", Lectures Notes in Computer Science, Springer-Verlag, 2008, vol. 5195, p. 347–362.
- [6] D. CACHERA, T. JENSEN, A. JOBIN, F. KIRCHNER. Inference of polynomial invariants for imperative programs: a farewell to Gro bner bases, in "Science of Computer Programming", 2014, vol. 93, 21 [DOI: 10.1016/J.SCICO.2014.02.028], https://hal.inria.fr/hal-00932351.
- [7] D. CACHERA, T. JENSEN, D. PICHARDIE, V. RUSU. *Extracting a Data Flow Analyser in Constructive Logic*, in "Theoretical Computer Science", 2005, vol. 342, n<sup>o</sup> 1, p. 56–78.
- [8] D. DEMANGE, V. LAPORTE, L. ZHAO, D. PICHARDIE, S. JAGANNATHAN, J. VITEK. Plan B: A Buffered Memory Model for Java, in "Proc. of the 40th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2013", Rome, Italy, ACM, 2013, http://hal.inria.fr/hal-00924716.
- [9] T. GENET, V. RUSU. Equational Approximations for Tree Automata Completion, in "Journal of Symbolic Computation", 2010, vol. 45(5):574-597, May 2010, n<sup>o</sup> 5, p. 574-597.
- [10] L. HUBERT, T. JENSEN, V. MONFORT, D. PICHARDIE. Enforcing Secure Object Initialization in Java, in "15th European Symposium on Research in Computer Security (ESORICS)", Lecture Notes in Computer Science, Springer, 2010, vol. 6345, p. 101-115.
[11] J.-H. JOURDAN, V. LAPORTE, S. BLAZY, X. LEROY, D. PICHARDIE. A formally-verified C static analyzer, in "POPL 2015: 42nd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages", Mumbai, India, ACM, January 2015, p. 247-259 [DOI: 10.1145/2676726.2676966], https://hal.inria.fr/ hal-01078386.

#### **Publications of the year**

#### **Doctoral Dissertations and Habilitation Theses**

- [12] O. F. ANDREESCU. Static Analysis of Functional Programs with an Application to the Frame Problem in Deductive Verification, Rennes 1, May 2017, https://hal.inria.fr/tel-01677897.
- [13] O. F. ANDREESCU. Static analysis of functional programs with an application to the frame problem in deductive verification, Université Rennes 1, May 2017, https://tel.archives-ouvertes.fr/tel-01682503.
- [14] P. BOLIGNANO. *Formal models and verification of memory management in a hypervisor*, Université Rennes 1, May 2017, https://tel.archives-ouvertes.fr/tel-01637937.
- [15] D. BÜHLER. *Structuring an Abstract Interpreter through Value and State Abstractions: EVA, an Evolved Value Analysis for Frama-C*, Université de Rennes 1, March 2017, https://hal.archives-ouvertes.fr/tel-01664726.
- [16] Y. ZAKOWSKI. Verification of a Concurrent Garbage Collector, École Normale Supérieur de Rennes, December 2017, https://hal.inria.fr/tel-01680213.

#### **Articles in International Peer-Reviewed Journal**

- [17] F. BESSON, S. BLAZY, P. WILKE.A Verified CompCert Front-End for a Memory Model Supporting Pointer Arithmetic and Uninitialised Data, in "Journal of Automated Reasoning", 2017, p. 1-48 [DOI: 10.1007/s10817-017-9439-Z], https://hal.inria.fr/hal-01656895.
- [18] H. CIRSTEA, S. LENGLET, P.-E. MOREAU. Faithful (Meta-)Encodings Of Programmable Strategies Into Term Rewriting Systems, in "Logical Methods in Computer Science", November 2017, vol. 13, n<sup>o</sup> 4, p. 1-54, Long version of the corresponding RTA-TLCA 15 paper [DOI: 10.23638/LMCS-13(4:16)2017], https:// hal.inria.fr/hal-01479030.
- [19] T. GENET, Y. SALMON. *Reachability Analysis of Innermost Rewriting extended version*, in "Logical Methods in Computer Science", 2017, https://hal.inria.fr/hal-01532090.
- [20] F. HONSELL, L. LIQUORI, P. MAKSIMOVIC, I. SCAGNETTO.LLFP : A Logical Framework for modeling External Evidence, Side Conditions, and Proof Irrelevance using Monads, in "Logical Methods in Computer Science", February 2017, https://arxiv.org/abs/1702.07214, https://hal.inria.fr/hal-01146059.

#### **Invited Conferences**

- [21] T. JENSEN. *Formal methods for software security (invited talk)*, in "FMF 2017 Forum "Méthodes Formelles"", Toulouse, France, January 2017, p. 1-61, https://hal.inria.fr/hal-01658549.
- [22] T. JENSEN. Formal methods for software security (invited talk), in "Journées Nationales 2017 Pré-GDR Sécurité Informatique", Paris, France, June 2017, p. 1-31, https://hal.inria.fr/hal-01658835.

[23] T. JENSEN. Hybrid information flow analysis against web tracking (invited talk), in "CRiSIS 2017 - 12th International Conference on Risks and Security of Internet and Systems", Dinard, France, September 2017, p. 1-33, https://hal.inria.fr/hal-01658896.

#### **International Conferences with Proceedings**

- [24] G. BARTHE, S. BLAZY, V. LAPORTE, D. PICHARDIE, A. TRIEU. Verified Translation Validation of Static Analyses, in "Computer Security Foundations Symposium", Santa-Barbara, United States, 30th IEEE Computer Security Foundations Symposium, August 2017, https://hal.inria.fr/hal-01588422.
- [25] F. BESSON, S. BLAZY, P. WILKE. CompCertS: A Memory-Aware Verified C Compiler using Pointer as Integer Semantics, in "ITP 2017 - 8th International Conference on Interactive Theorem Proving", Brasilia, Brazil, ITP 2017: Interactive Theorem Proving, Springer, September 2017, vol. 10499, p. 81-97 [DOI : 10.1007/978-3-319-66107-0\_6], https://hal.inria.fr/hal-01656875.
- [26] M. BIERNACKA, D. BIERNACKI, S. LENGLET, P. POLESIUK, D. POUS, A. SCHMITT. Fully Abstract Encodings of λ-Calculus in HOcore through Abstract Machines, in "LICS 2017", Reykjavik, Iceland, Proceedings of LICS 2017, June 2017, To appear, https://hal.inria.fr/hal-01479035.
- [27] S. BLAZY, D. PICHARDIE, A. TRIEU. Verifying Constant-Time Implementations by Abstract Interpretation, in "European Symposium on Research in Computer Security", Oslo, Norway, 22nd European Symposium on Research in Computer Security, September 2017, https://hal.inria.fr/hal-01588444.
- [28] D. KÄSTNER, J. BARRHO, U. WÜNSCHE, M. SCHLICKLING, B. SCHOMMER, M. SCHMIDT, C. FER-DINAND, X. LEROY, S. BLAZY. CompCert: Practical Experience on Integrating and Qualifying a Formally Verified Optimizing Compiler, in "ERTS2 2018 - Embedded Real Time Software and Systems", Toulouse, France, 3AF, SEE, SIE, January 2018, https://hal.inria.fr/hal-01643290.
- [29] D. KÄSTNER, X. LEROY, S. BLAZY, B. SCHOMMER, M. SCHMIDT, C. FERDINAND. Closing the Gap – The Formally Verified Optimizing Compiler CompCert, in "SSS'17: Safety-critical Systems Symposium 2017", Bristol, United Kingdom, Developments in System Safety Engineering: Proceedings of the Twentyfifth Safety-critical Systems Symposium, CreateSpace, February 2017, p. 163-180, https://hal.inria.fr/hal-01399482.
- [30] S. LENGLET, A. SCHMITT.HOπ in Coq, in "CPP 2018 The 7th ACM SIGPLAN International Conference on Certified Programs and Proofs", Los Angeles, United States, January 2018, 14 [DOI: 10.1145/3167083], https://hal.inria.fr/hal-01614987.
- [31] M. LESLOUS, V. VIET TRIEM TONG, J.-F. LALANDE, T. GENET. GPFinder: Tracking the Invisible in Android Malware, in "12th International Conference on Malicious and Unwanted Software", Fajardo, Puerto Rico, IEEE Conputer Society, October 2017, https://hal-centralesupelec.archives-ouvertes.fr/hal-01584989.
- [32] Y. ZAKOWSKI, D. CACHERA, D. DEMANGE, G. PETRI, D. PICHARDIE, S. JAGANNATHAN, J. VITEK. Verifying a Concurrent Garbage Collector using a Rely-Guarantee Methodology, in "ITP 2017 8th International Conference on Interactive Theorem Proving", Brasília, Brazil, Lecture Notes in Computer Science, Springer, September 2017, vol. 10499, p. 496-513 [DOI : 10.1007/978-3-319-66107-0\_31], https://hal.inria.fr/hal-01613389.

[33] Y. ZAKOWSKI, D. CACHERA, D. DEMANGE, D. PICHARDIE. Verified Compilation of Linearizable Data Structures: Mechanizing Rely Guarantee for Semantic Refinement, in "SAC 2018 - The 33rd ACM/SIGAPP Symposium On Applied Computing", Pau, France, April 2018, p. 1-10, https://hal.archives-ouvertes.fr/hal-01653620.

#### **National Conferences with Proceeding**

[34] G. CABON, A. SCHMITT.Non-Interference through Annotated Multisemantics, in "28ièmes Journées Francophones des Langages Applicatifs", Gourette, France, January 2017, https://hal.archives-ouvertes.fr/hal-01503094.

#### **Conferences without Proceedings**

- [35] G. CABON, A. SCHMITT. Annotated multisemantics to prove Non-Interference analyses, in "PLAS 2017 -ACM SIGSAC Workshop on Programming Languages and Analysis for Security", Dallas, United States, PLAS '17 Proceedings of the 2017 Workshop on Programming Languages and Analysis for Security, ACM, October 2017, p. 49-62 [DOI: 10.1145/3139337.3139344], https://hal.archives-ouvertes.fr/hal-01656404.
- [36] F. SAUDEL, S. BLAZY, F. BESSON. Confusion de Type en C++: État de l'Art et Difficultés de Détection, in "RESSI 2017 - Rendez-vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information", Grenoble/Autrans, France, May 2017, p. 1-5, https://hal.inria.fr/hal-01656979.

#### **Research Reports**

- [37] M. BIERNACKA, D. BIERNACKI, S. LENGLET, P. POLESIUK, D. POUS, A. SCHMITT.Fully Abstract Encodings of λ-Calculus in HOcore through Abstract Machines, Inria, April 2017, n<sup>0</sup> RR-9052, https://hal. inria.fr/hal-01507625.
- [38] T. GENET.A Short Isabelle/HOL Tutorial for the Functional Programmer, IRISA, 2017, https://hal.inria.fr/ hal-01208577.
- [39] T. GENET. Automata Completion and Regularity Preservation, IRISA, Inria Rennes, April 2017, https://hal.archives-ouvertes.fr/hal-01501744.
- [40] T. GENET, T. HAUDEBOURG, T. JENSEN. Verifying Higher-Order Functions with Tree Automata: Extended Version, Irisa, October 2017, p. 1-20, https://hal.inria.fr/hal-01614380.
- [41] Y. ZAKOWSKI, D. CACHERA, D. DEMANGE, D. PICHARDIE. Compilation of Linearizable Data Structures: A Mechanised RG Logic for Semantic Refinement, ENS Rennes; IRISA, Inria Rennes; Université Rennes 1, June 2017, https://hal.archives-ouvertes.fr/hal-01538128.

#### **Other Publications**

[42] T. GENET.SPAN+AVISPA for Verifying Cryptographic Protocols, 2017, This is a video tutorial to learn how to use SPAN+AVISPA to automatically check security properties on cryptographic protocols, https://hal.inria. fr/hal-01532086.

## **Project-Team CIDRE**

# Confidentialité, Intégrité, Disponibilité et Répartition

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: CNRS CentraleSupélec Université Rennes 1

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Security and Confidentiality

### **Table of contents**

1.	Personnel	. 151
2.	Overall Objectives	. 152
3.	Research Program	. 152
	3.1. Our perspective	152
	3.2. Intrusion Detection / Security Events Monitoring and Management	153
	3.3. Privacy	154
4.	Application Domains	155
5.	Highlights of the Year	. 155
6.	New Software and Platforms	. 156
	6.1. Blare	156
	6.2. GNG	157
	6.3. GroddDroid	157
	6.4. Kharon	157
	6.5. StarLord	158
	6.6. SpecCert	158
	6.7. HardBlare	158
	6.8. Conductor	159
	6.9. Platforms	159
7.	New Results	. 159
	7.1. Intrusion Detection	159
	7.1.1. Intrusion Detection in Distributed Systems	159
	7.1.2. Illegal Information Flow Detection	160
	7.1.3. Intrusion Detection in Low-Level Software Components	162
	7.1.4. Vizualization	163
	7.2. Privacy	163
	7.3. Security of Communicating and Distributed Systems	164
	7.3.1. Routing Protocol for Tactical Mobile Ad Hoc Networks	164
	7.3.2. Decentralized Cryptocurrency Systems	164
	7.3.3. Large Scale Systems	165
8.	Bilateral Contracts and Grants with Industry	. 165
	8.1. Bilateral Contracts with Industry	165
	8.2. Bilateral Grants with Industry	166
9.	Partnerships and Cooperations	. 167
	9.1. Regional Initiatives	167
	9.2. National Initiatives	168
	9.3. International Initiatives	169
	9.4. International Research Visitors	169
10.	Dissemination	. 169
	10.1. Promoting Scientific Activities	169
	10.1.1. Scientific Events Organisation	169
	10.1.1.1. General Chair, Scientific Chair	169
	10.1.1.2. Member of the Organizing Committees	170
	10.1.2. Scientific Events Selection	170
	10.1.2.1. Member of the Conference Program Committees	170
	10.1.2.2. Reviewer	170
	10.1.3. Journal	171
	10.1.3.1. Member of the Editorial Boards	171
	10.1.3.2. Reviewer - Reviewing Activities	171
	10.1.4. Invited Talks	171

10.1.5. Leadership within the Scientific Community	171
10.1.6. Research Administration	171
10.2. Teaching - Supervision - Juries	172
10.2.1. Certification	172
10.2.2. Teaching	172
10.2.3. Supervision	176
10.2.3.1. Theses defended in 2017	176
10.2.3.2. Theses in progress	177
10.2.3.3. Supervision of external PhD candidates	177
10.2.4. Juries	178
10.3. Popularization	178
11. Bibliography	178

#### **Project-Team CIDRE**

Creation of the Project-Team: 2011 July 01

#### **Keywords:**

#### **Computer Science and Digital Science:**

A1.2.8. - Network security
A1.3. - Distributed Systems
A3.3.1. - On-line analytical processing
A3.5.2. - Recommendation systems
A4.1.1. - Malware analysis
A4.1.2. - Hardware attacks
A4.4. - Security of equipment and software
A4.8. - Privacy-enhancing technologies
A4.9.1. - Intrusion detection
A4.9.2. - Alert correlation
A7.1. - Algorithms

#### **Other Research Topics and Application Domains:**

B6.5. - Information systems B9.8. - Privacy

## 1. Personnel

#### **Research Scientists**

Emmanuelle Anceaume [CNRS, Researcher] Michel Hurfin [Inria, Researcher, HDR] Mohamed Kasraoui [Univ de Rennes I, Researcher, until Aug 2017]

#### **Faculty Members**

Christophe Bidan [Team leader, CentraleSupélec, Professor, HDR] Gilles Guette [Univ de Rennes I, Associate Professor] Guillaume Hiet [CentraleSupélec, Associate Professor] Jean-François Lalande [CentraleSupélec, Associate Professor, from Sep 2017, HDR] Ludovic Mé [CentraleSupélec, Professor, HDR] Guillaume Piolle [CentraleSupélec, Associate Professor] Eric Totel [CentraleSupélec, Professor, HDR] Frédéric Tronel [CentraleSupélec, Associate Professor] Valérie Viet Triem Tong [CentraleSupélec, Associate Professor, HDR]

#### **External Collaborators**

Frédéric Majorczyk [DGA] Sébastien Gambs [UQAM]

#### **Technical Staff**

Antoine Guellier [CNRS, from Jun 2017] Christopher Humphries [Inria, until Aug 2017]

#### **PhD Students**

Solenn Brunet [Orange Labs, until Oct 2017] Vasile Cazacu [CNRS, from Feb 2017] Ronny Chevalier [Hewlet Packard France] Kun He [IMT Atlantique, until Sep 2017] Deepak Subramanian [IMT Atlantique, until Sep 2017] Damien Crémilleux [CentraleSupélec] Aurélien Dupin [Thales] Laurent Georget [Univ de Rennes I, until Sep 2017] Florian Grandhomme [Univ de Rennes I, until Sep 2017] Pierre Graux [Inria, from Oct 2017] David Lanoé [Inria] Laetitia Leichtnam [Ministère de la Défense] Mourad Leslous [Inria] Thomas Letan [ANSSI, until Sep 2017] Pernelle Mensah [Bell Labs (Nokia)] Mounir Nasr Allah [CentraleSupélec] Aurélien Trulla [Inria] Charles Arya Xosanavongsa [Thales]

#### Administrative Assistant

Lydie Mabil [Inria]

## 2. Overall Objectives

#### 2.1. CIDRE in Brief

Our long term ambition is to contribute to the building of distributed systems that are trustworthy and respectful of privacy, even when some nodes in the system have been compromised.

With this objective in mind, the CIDRE team focuses mainly on the two following topics: Intrusion Detection and Privacy Protection.

## 3. Research Program

#### **3.1. Our perspective**

For many aspects of our everyday life, we heavily rely on information systems, many of which are based on massively networked devices that support a population of interacting and cooperating entities. While these information systems become increasingly open and complex, accidental and intentional failures get considerably more frequent and severe.

Two research communities traditionally address the concern of accidental and intentional failures: the distributed computing community and the security community. While both communities are interested in the construction of systems that are correct and secure, an ideological gap and a lack of communication exist between them that is often explained by the incompatibility of the assumptions each of them traditionally makes. Furthermore, in terms of objectives, the distributed computing community has favored systems availability while the security community has focused on integrity and confidentiality, and more recently on privacy. Our long term ambition is to contribute to the building of distributed systems that are trustworthy and respectful of privacy, even when some nodes <sup>0</sup> in the system have been compromised. For that purpose, we are convinced that combining classical security approaches and distributed computing paradigms is an interesting way to enforce the security of large-scale distributed systems. More specifically, since a distributed system is composed of nodes, we assert that the security of large-scale distributed systems has to be addressed at three complementary levels:

- the level of each node: each standalone node has to enforce its own security;
- the level of an *identified* set of *trusted* nodes: the *trusted* nodes can *collaborate* to enforce together their security;
- the level of fully open large-scale distributed and dynamic systems: distributed computing paradigms such as consensus algorithms can be applied to cope with the possible presence of malicious nodes.

Notice that using a distributed architecture can also be an approach allowing the nodes to enforce their security without the need of a trusted third party.

The research activities of the CIDRE project-team focus mainly on the two following research axis:

- Intrusion Detection System: the objective is to detect any suspicious events with regard to the security by analyzing some data generated on the monitored system.
- **Privacy-preserving Services:** the objective is to ensure users' privacy even when this property seems incompatible with the provided services, like social networks or location-based services.

In all our studies, we consider a priori that the attacker is omnipotent. He can acts as he wants. Nevertheless, since our team is not specialized in cryptography, we consider that we can rely on strong unbroken cryptosystems.

#### 3.2. Intrusion Detection / Security Events Monitoring and Management

Today, we have not yet fully entered into a world of "security by design". Security remains often a property that is considered a posteriori, when the system is deployed, which often results in applying patches when vulnerabilities are discovered (also called a "patch and pray" approach). Unfortunately, despite patching, the number of vulnerabilities remains high, as evidenced by the number of vulnerabilities published each year in the Common Vulnerabilities and Exposures (CVE) system. Thus, it is important to be able to early detect cyber-attacks, especially when they exploit vulnerabilities that are unknown. However, the efficiency of security events monitoring and management systems (including the IDS - Intrusion Detection Systems) is still an open issue today. Indeed, they are often unable to effectively deal with huge numbers of security events, and they usually produce too many false alarms yet missing some attacks. So one of the main research challenges in IT security remains the definition of efficient security events monitoring systems, i.e., that enable both to process a huge number of security events and to detect any attacks without flooding the security analysts with false alarms.

By exploiting vulnerabilities in operating systems, applications, or network services, an attacker can defeat preventive security mechanisms and violate the security policy of the whole system. The goal of an Intrusion Detection Systems (IDS) is to detect such violations by analyzing some *security events* generated on a monitored system. Ideally, the IDS should produce an alert for any violation (no *false negative*), and only for violations (no *false positive*).

To produce alerts, two detection techniques exist: the misuse based detection and the anomaly based detection. A misuse based detection is actually a signature based detection approach : it allows to detect only the attacks whose signature is available. From our point of view, while useful in practice, misuse detection is intrinsically limited. Indeed, it requires to update in real-time the database of signatures, similarly to what has to be done for antivirus tools. The CIDRE project-team follows the alternative approach, namely the anomaly approach, which consists in detecting a deviation from a referenced behavior. Our contributions on anomaly-based IDS follow three axis:

<sup>&</sup>lt;sup>0</sup>The term node either refers to a device that hosts a network client or service or to the process that runs this client or service.

- **Illegal Information Flow Detection:** our goal is to detect information flows in the monitored system (either a node or a set of trusted nodes) that are allowed by the access control mechanism, but are illegal from the security policy point of view. This approach is particularly appealing to detect intrusions in a standalone node, such as a smartphone.
- Anomaly-Based Detection in Distributed Applications: our goal is to specify the normal behavior based on either a formal specification of the distributed application, or previous executions. This approach is particularly appealing to detect intrusions in industrial control systems since these systems exhibit well-defined behaviors at different levels: network level (network communication patterns, protocol specifications, etc.), control level (continue and discrete process control laws), or even the state of the local resources (memory or CPU).
- **Online data analytics:** our goal is to estimate on the fly different statistics or metrics on distributed input streams to detect abnormal behavior with respect to a well-defined criterion such as the distance between different streams, their correlation or their entropy.

Beside the anomaly-based IDS, we have also led research work on alert correlation and visualisation of security events. Indeed, in large systems, multiple (host and network) IDS and many sensors are deployed and they continuously and independently generate notifications (event's observations, warnings and alerts). To cope with this huge amount of collected data, we have studied two different approaches, each with specific goal:

- Alert Correlation System: the alerts of *low level* IDSes can be viewed as *security events* of a *high level* IDS whose goal is to correlate these alerts. An alert correlation system aims at exploiting the known relationships between some elements that appear in the flow of low level notifications to generate high semantic meta-alerts. The main goal is to reduce the number of alerts (and especially, false positive) returned to the security analysts and to allow a higher level analysis of the situation (situational awareness).
- Visualization Tools: a visualization tools aims at relying on the capacity of human beings to detect patterns and outliers in datasets when these datasets are properly visually represented. Human beings also know pieces of contextual information that are very difficult to formalize so as to make them usable by a computer. Visualization is therefore a very useful complementary tool to detect abnormal events in real time (monitoring), to search for malicious events in log files (data exploration and forensics) and to communicate results (reporting).

#### 3.3. Privacy

In a world of ubiquitous technologies, each individual constantly leaves digital traces related to his activities and interests. The current business plan of many web services such as social networks, is based on the sale of these digital traces. Of course, this is usually done in a legal way, the license of use clearly stating that the user gives the right to the service provider for using his personal data. However, on the one hand, users generally do not read these licenses, and on the other hand, these licenses are usually very vague on the use of personal data <sup>0</sup>. In addition these digital traces can potentially be stolen and maliciously used, they must therefore be protected. In this context, users' privacy is now recognized as a fundamental individual right. Any new IT service should thus follow the *privacy-by-design* approach: privacy issues have to be studied from the earliest phase of a project by taking into account the multi-stakeholders and transdisciplinary aspects in order to ensure proper, end-to-end private data protection properties.

In the CIDRE project, we mainly focus on domains in which privacy issues collide with provided services. Here are some concrete examples of such domains:

- **Location-based services:** the challenge is to design services that depend on the user's location while preserving the privacy of his location;
- **Social networks:** the challenge is to demonstrate that it is possible to design social networks respectful of users' privacy;

<sup>&</sup>lt;sup>0</sup>Besides, it has been shown that service providers do not necessarily comply with their own license.

- **Mobile services:** given that such services are based on user's identity, the challenge is to design mobile services while preserving the users' anonymity;
- Ad-hoc netwoks: in ad-hoc networks, any participant can potentially know the relative location of the other participants. Thus, the issue is to allow nodes to forward messages while preserving the privacy of the communications.

For all of these domains, we have proposed new Privacy-Enhancing Techniques (PETs) based on a mix of different foundations such as cryptographic techniques, security policies and access control mechanisms, just to name a few. More generally, we think that a major option to protect users' privacy consists in using a decentralized architecture that enables to transfer control and services from the service providers to the users.

The concept of IDS seems to be in contradiction with the users' privacy. Indeed, an IDS is a monitoring system that needs to collect and analyze information coming from different levels such as network, applications and OS, this information being able to include users' personal data. However, we are confident that IDS and privacy are not completely antagonist. In particular, integrating some privacy features inside an IDS to build a privacy-preserving IDS may allow to limit the amount of information that can leak if one of the nodes within the system is compromised. On the other hand, enabling IDS to detect attacks against privacy as well as security violations can extend the range of their applicability.

## 4. Application Domains

#### 4.1. Security is Required Everywhere

With the infiltration of computers and software in almost all aspects of our modern life, security can nowadays be seen as an absolutely general concern. As such, the results of the research targeted by CIDRE apply to a wide range of domains. It is clear that critical systems, in which security (and safety) is a major concern, can benefit from ideas such as dynamic security policy monitoring. On the other hand, systems used by the general public (basically, the internet and services such as web or cloud services, social networks, location-based services, etc.) can also benefit from results obtained by CIDRE, in particular to solve some of the privacy issues raised by these systems that manipulate huge amount of personal data. In addition, systems are getting more and more complex, decentralized, distributed, or spontaneous. Cloud computing, in particular, brings many challenges that could benefit from ideas, approaches and solutions studied by CIDRE in the context of distributed systems.

Industrial Control Systems (ICS) and in particular Supervisory Control and Data Acquisition are also new application domains for intrusion detection. The Stuxnet attack has emphasized the vulnerability of such critical systems which are not totally isolated anymore. Securing ICS is challenging since modifications of the systems, for example to patch them, are often not possible. High availability requirements also often conflict with preventive approaches. In this case, security monitoring is appealing to protect such systems against malicious activities. Intrusion detection in ICS is not fundamentally different from traditional approaches. However, new hypotheses and constraints need to be taken into account, which also bring interesting new research challenges.

## 5. Highlights of the Year

#### 5.1. Highlights

This year, the CIDRE team would like to emphasize the following publications that appeared in major academic venues:

• Formal verification of an information flow monitor, presented at SEFM'17 [11]. See below (5.1.1) for a more complete description of this work.

- Automated quantitative information flow analysis for imperative deterministic programs, presented at POPL'17 [8].
- Reconstruction of connectivity graph for cloud infrastructures, presented at NCA'2017 [17]
- Co-processor-based Behavior Monitoring: Application to the Detection of Attacks Against the SMM, presented at ACSAC'17 [10]

#### 5.1.1. Awards

Laurent Georget, Mathieu Jaume (LIP6), Guillaume Piolle, Frédéric Tronel and Valérie Viet Triem Tong received the best paper award at the SEFM'17 conference, which is a well established conference focused on the link between software development and formal methods. This publication is based on the work realized by Laurent Georget during his PhD. It focuses on the automated verification of the correctness of an information flow monitor that operates at the kernel level (Linux kernel). This information flow monitor relies on the Linux Security Module (LSM hereafter) framework. This framework has been designed for mandatory access control. This work tries to answer the question of its correctness when used for information flow monitoring. The verification is operated by a GCC plugin during the compilation phase of a full Linux kernel. Based on an ad-hoc static analysis, it can determine if the LSM hooks are correctly placed with respect to a property of complete mediation of systems calls. Each system call that is known to generate an information flow during its execution (34 system calls on a grand total of 340) is analyzed to determine if the LSM framework through the hooks it provides can intercept each execution that potentially generates an information flow. We have demonstrated that for 4 system calls, the hooks are not well placed, and discovered that 4 systems calls are simply lacking LSM hooks. A patch has been produced to improve this situation.

**BEST PAPERS AWARDS :** 

[11] 15th International Conference on Software Engineering and Formal Methods (SEFM 2017). L. GEORGET, M. JAUME, G. PIOLLE, F. TRONEL, V. VIET TRIEM TONG.

## 6. New Software and Platforms

#### 6.1. Blare

To detect intrusion using information flows

KEYWORDS: Cybersecurity - Intrusion Detection Systems (IDS) - Data Leakage Protection SCIENTIFIC DESCRIPTION: Blare implements our approach of illegal information flow detection for a single node (Android and Linux kernel, JVM) and a set of nodes (monitoring of flows between linux machines). FUNCTIONAL DESCRIPTION: Blare IDS is a set of tools that implements our approach to illegal information flow detection for a single node and a set of nodes.

NEWS OF THE YEAR: During this year, Laurent Georget has modified the implementation of Blare in order to correctly monitor the kernel system calls with LSM hooks. He add also ported this new version of Blare to the Lollipop Android emulator.

- Partner: CentraleSupélec
- Contact: Frédéric Tronel
- Publications: Information Flow Tracking for Linux Handling Concurrent System Calls and Shared Memory - Verifying the Reliability of Operating System-Level Information Flow Control Systems in Linux - Monitoring both OS and program level information flows to detect intrusions against network servers - Experimenting a Policy-Based HIDS Based on an Information Flow Control Model - Introducing reference flow control for intrusion detection at the OS level - Blare Tools: A Policy-Based Intrusion Detection System Automatically Set by the Security Policy - Diagnosing intrusions in Android operating system using system flow graph - Intrusion detection in distributed systems, an approach based on taint marking - BSPL: A Language to Specify and Compose Fine-grained Information Flow Policies - Information Flow Policies vs Malware - A taint marking approach to confidentiality violation detection - Designing information flow policies for Android's operating

system - Information Flow Control for Intrusion Detection derived from MAC Policy - Flow based interpretation of access control: Detection of illegal information flows - A taint marking approach to confidentiality violation detection

• URL: http://www.blare-ids.org/

#### 6.2. GNG

Security Supervision by Alert Correlation

KEYWORDS: Intrusion Detection Systems (IDS) - SIEM

SCIENTIFIC DESCRIPTION: GNG is an intrusion detection system that correlates different sources (such as different logs) in order to identify attacks against the system. The attack scenarios are defined using the Attack Description Langage (ADeLe) proposed by our team, and are internally translated to attack recognition automatons. GNG intends to define time efficient algorithms based on these automatons to recognize complex attack scenarios.

- Partner: CentraleSupélec
- Contact: Eric Totel
- Publication: A Language Driven Intrusion Detection System for Events and Alerts Correlation
- URL: http://www.rennes.supelec.fr/ren/perso/etotel/GNG/index.html

#### 6.3. GroddDroid

KEYWORDS: Android - Detection - Malware

SCIENTIFIC DESCRIPTION: GroddDroid automates the dynamic analysis of a malware. When a piece of suspicious code is detected, groddDroid interacts with the user interface and eventually forces the execution of the identified code. Using Blare (Information Flow Monitor), GroddDroid monitors how an execution contaminates the operating system. The output of GroddDroid can be visualized in an web browser. GroddDroid is used by the Kharon software.

FUNCTIONAL DESCRIPTION: GroddDroid 1 - locates suspicious code in Android application 2 - computes execution paths towards suspicious code 3 - forces executions of suspicious code 4 - automate the execution of a malware or a regular Android application

NEWS OF THE YEAR: In 2017, GroddDroid has integrated the work of Mourad Leslous, who have implemented GFinder. GPFinder improves the computation of control flow paths by taking into account the Android framework. The end of the year has been used to clean the code and to improves the graphical interface.

- Partners: CentraleSupélec Insa Centre Val-de-Loire
- Contact: Valérie Viet Triem Tong
- Publications: Kharon dataset: Android malware under a microscope GroddDroid: a Gorilla for Triggering Malicious Behaviors - GPFinder: Tracking the Invisible in Android Malware - Information flows at OS level unmask sophisticated Android malware
- URL: http://kharon.gforge.inria.fr/grodddroid.html

#### 6.4. Kharon

KEYWORDS: Android - Malware - Dynamic Analysis

FUNCTIONAL DESCRIPTION: Kharon is a software for managing Android application analysis. Kharon uses the results of the GroddDroid software. The user can submit one or several applications to Kharon and get a graph of the innformation flows that occured at system level and that have been caused by the application.

Kharon is used in the Kharon platform for the analysis of malicious applications. This platform is deployed at the high security laboratory (LHS) of Rennes.

- Author: Sébastien Campion
- Partners: CentraleSupélec Insa Centre Val-de-Loire
- Contact: Valérie Viet Triem Tong
- URL: http://kharon.gforge.inria.fr/

#### 6.5. StarLord

#### KEYWORDS: Security - SIEM

FUNCTIONAL DESCRIPTION: In the domain of security event visualisation, we have developed a prototype called StarLord. Basically, this software is able to parse heterogeneous logs, and to extract from each line of logs a set of security objects. Moreover, some of these objects appears in several lines of different logs. These lines are thus linked by the sharing of one or more security objects. When we analyse the lines of logs, we are thus able to generate graphs that represents the links between the different objects discovered in the logs. These graphs are thus displayed in 3D in order for the administrator to investigate easily the relations between the logs and the relations between the logs and some particular indicators of compromission. The tool permits to discover visually the activity of an attacker on the supervised system.

- Authors: Ludovic Mé, Eric Totel, Nicolas Prigent and Laetitia Leichtnam
- Contact: Eric Totel
- Publication: STARLORD: Linked Security Data Exploration in a 3D Graph

#### 6.6. SpecCert

KEYWORDS: Formal methods - Coq

FUNCTIONAL DESCRIPTION: SpecCert is a framework for specifying and verifying Hardware-based Security Enforcement (HSE) mechanisms against hardware architecture models. HSE mechanisms form a class of security enforcement mechanism such that a set of trusted software components relies on hardware functions to enforce a security policy.

- Participant: Thomas Letan
- Partners: ANSSI CentraleSupélec
- Contact: Guillaume Hiet
- Publications: SpecCert: Specifying and Verifying Hardware-based Security Enforcement SpecCert: Specifying and Verifying Hardware-based Software Enforcement
- URL: https://github.com/lethom/speccert

#### 6.7. HardBlare

KEYWORDS: Intrusion Detection Systems (IDS) - FPGA - Static analysis

FUNCTIONAL DESCRIPTION: HardBlare is a hardware/software framework to implement hardware DIFC on Xilinx Zynq Platform. HardBlare consists of three components : 1) the VHDL code of the coprocessor, 2) a modified LLVM compiler to compute the static analysis, and 3) a dedicated Linux kernel. This last component is a specific version of the Blare monitor.

- Partners: CentraleSupélec Lab-STICC
- Contact: Guillaume Hiet
- Publications: ARMHEx: A hardware extension for DIFT on ARM-based SoCs ARMHEx: a framework for efficient DIFT in real-world SoCs ARMHEx: embedded security through hardware-enhanced information flow tracking HardBlare: a Hardware-Assisted Approach for Dynamic Information Flow Tracking A portable approach for SoC-based Dynamic Information Flow Tracking implementations Towards a hardware-assisted information flow tracking ecosystem for ARM processors HardBlare: an efficient hardware-assisted DIFC for non-modified embedded processors

#### 6.8. Conductor

KEYWORDS: Intrusion Detection Systems (IDS) - Static analysis - Instrumentation

FUNCTIONAL DESCRIPTION: Conductor contains three main components: a static analysis to extract the expected behavior of the target, an instrumentation module to add instructions to the target's code in order to send messages to the co-processor, and an intrusion detection engine executed on the co-processor. The latter processes the messages sent by the instrumented target, describing its current behavior. This behavior is then compared against the expected behavior previously extracted by the static analysis.

- Participants: Ronny Chevalier, Guillaume Hiet, Maugan Villatel and David Plaquin
- Partners: CentraleSupélec HP Labs
- Contact: Ronny Chevalier
- Publication: Co-processor-based Behavior Monitoring: Application to the Detection of Attacks Against the System Management Mode

#### 6.9. Platforms

#### 6.9.1. Kharon platform

The Kharon platform is under development in the LHS of Rennes and should be ready to use in the beginning of 2018. This experimental platform aims to analyze Android malware using a set of software developed by the CIDRE team. Software that are involved are:

- The Blare IDS http://www.blare-ids.org/, and in particular the AndroBlare version, for tracking information flows of malware;
- The GroddDroid software http://kharon.gforge.inria.fr/grodddroid.html, for manipulating the malware statically and dynamically;
- The GPFinder software http://kharon.gforge.inria.fr/gpfinder.html, for computing paths in the malware's control flow;
- The kharon software that handles the orchestration of a bunch of malware, the server and a set of smartphones.

The Kharon platform will be used for analysing malware as soon as they appear in the wild. The analysis results will be stored for further experiments and statistics.

## 7. New Results

#### 7.1. Intrusion Detection

#### 7.1.1. Intrusion Detection in Distributed Systems

Alert Correlation: In large systems, multiple (host and network) Intrusion Detection Systems (IDS) and many sensors are usually deployed. They continuously and independently generate notifications (event's observations, warnings and alerts). To cope with this amount of collected data, alert correlation systems have to be designed. An alert correlation system aims at exploiting the known relationships between some elements that appear in the flow of low level notifications to generate high semantic meta-alerts. The main goal is to reduce the number of alerts returned to the security administrator and to allow a higher level analysis of the situation. However, producing correlation rules is a highly difficult operation, as it requires both the knowledge of an attacker, and the knowledge of the functionalities of all IDSes involved in the detection process. In the context of the PhD of Erwan Godefroy, we focus on the transformation process that allows to translate the description of a complex attack scenario into correlation rules and its assessment. We show that, once a human expert has provided an action tree derived from an attack tree, a fully automated transformation process can generate exhaustive correlation rules that would be tedious and error prone to enumerate by hand. This is a topdown approach to correlation rule generation. With the PhD of Charles Xosanavongsa, we tackle the problem of a bottom-up approach that consists in discovering automatically the events or alerts that have been produced by the attacker activity. The objective is to classify automatically all suspicious entries in heterogeneous logs relative to a given attack. This requires to exhibit all log entries that are causally linked, and permits to produce a correlation rule that could detect later a new occurence of the attack.

**Intrusion Detection in Cloud Infrastructure**: Prior to detecting intrusion, it can be useful to know how the supervised system is vulnerable to attacks. Such result is obtained during a risk analysis phase in usual systems. In the PhD thesis of Pernelle Mensah, we try to automate the generation of the description of all possible attacks against a Cloud infrastructure. This work is divided in two separate steps: (1) We first discover the topology of the virtual machines executing in the cloud infrastructure [16], [17] and (2) Build in a second phase a topological attack graph that represents all possible known attacks on the virtual infrastructure. This graph will be later used either to adapt counter-measures to known attacks, or to generate automatically correlation rules to detect the described attacks.

**Inferring the normal behavior of an application**: We propose an approach to detect intrusions that affect the behavior of distributed applications. To determine whether an observed behavior is normal or not (occurence of an attack), we rely on a model of normal behavior. This model has been built during an initial training phase (machine learning approach). During this preliminary phase, the application is executed several times in a safe environment. The gathered traces (sequences of actions) are used to generate an automaton that characterizes all these acceptable behaviors. To reduce the size of the automaton and to be able to accept more general behaviors that are close to the observed traces, the automaton is transformed. These transformations may lead to introduce unacceptable behaviors. Our current work solves this problem by characterizing the acceptable behaviors with invariant properties that they must verify. During the PhD thesis of David Lanoe, we enhanced the model building. Moreover, we assess this solution, by applying it to a distributed file system called XtreemFS. We show that it is possible to build the model of this given application, and to detect attack against XtreemFS, without producing too much false positives.

This approach is particularly appealing to detect intrusions in industrial control systems since these systems exhibit well-defined behaviors at different levels: network level (network communication patterns, protocol specifications, etc.), control level (continue and discrete process control laws), or even the state of the local resources (memory or CPU). Industrial control systems (ICS) can be subject to highly sophisticated attacks which may lead the process towards critical states. Due to the particular context of ICS, protection mechanisms are not always practical, nor sufficient. On the other hand, developing a process-aware intrusion detection solution with satisfactory alert characterization remains an open problem. Sophisticated process-aware attacks targeting industrial control systems require adequate detection measures taking into account the physical process. We propose an approach relying on automatically mined process specifications to detect attacks on sequential control systems. The specifications are synthesized as monitors that read the execution traces and report violations to the operator. In contrast to other approaches, a central aspect of our method consists in reducing the number of mined specifications suffering from redundancies. We evaluate our approach on a hardware-in-the-loop testbed with a complex physical process model and discuss our approach's mining efficiency and attack detection capabilities. This work has been submitted to the Safeprocess'18 conference.

#### 7.1.2. Illegal Information Flow Detection

Our research work on intrusion detection based on information flow has been initiated in 2002. This research work has resulted in Blare, a framework for Intrusion Detection Systems <sup>0</sup>, including KBlare, an implementation as a Linux Security Module (LSM), JBlare, an implementation for the Java Virtual Machine (JVM), and AndroBlare, for Android applications.

**Information Leaks**: Qualitative information flow aims at detecting information leaks, whereas the emerging quantitative techniques target the estimation of information leaks. Quantifying information flow in the presence of low inputs is challenging, since the traditional techniques of approximating and counting the reachable states of a program no longer suffice. We propose an automated quantitative information flow analysis for imperative deterministic programs with low inputs. The approach relies on a novel abstract domain, the cardinal abstraction, in order to compute a precise upper-bound over the maximum leakage of batch-job programs. We prove the soundness of the cardinal abstract domain by relying on the framework of abstract interpretation. We also prove its precision with respect to a flow-sensitive type system for the two-point security lattice. This approach has been published in POPL'17 [8].

<sup>&</sup>lt;sup>0</sup>http://www.blare-ids.org/

**Correct information flow monitoring by design**: As mentionned previously, our research team is developing an information monitor called **Blare**. Like most of its competitors (e.g. Laminar or Weir) our solution is based on the Linux Security Module (LSM) framework. However, this framework was initially designed with access control in mind. A natural question arises from this matter of fact: does the LSM framework can be used to correctly track information flow (at the operating system level) ? In the context of his PhD thesis, Laurent Georget has studied this very same question.

To tackle this problem, Laurent Georget has designed an ad hoc static analysis that run as a GCC plugin during the Linux kernel compilation. This analysis can prove (or disprove) the fact that LSM hooks within a chosen set of system calls (known to realize information flows between operating systems containers like files, sockets or pipe) are placed at correct locations so as to intercept these possible information flows. The experiments conducted by Laurent Georget have revealed that on an initial set of 38 system calls, 28 were correctly instrumented by LSM, 4 of them were equipped with a LSM hook that could miss some information flow (under certain circumstances), 3 were simply lacking a LSM hook, and 3 false positives had to be manually analyzed and requalified. Laurent Georget was able to produce a kernel patch to remove all missing and misplaced hooks. This patch can be prove to be correct using the same tool. This contribution was published at FormaliSE 2017 [12].

We had detected for a long time a subtle bug in our information flow monitor implementation (Blare) that we were able to track down to a race condition between two concurrent system calls reading and writing into the same pipe. Laurent Georget has proposed during its PhD an elegant solution to this complex problem: he proposed to divide each information flow into three stages: the activation, the execution and the deactivation. Only the activation and deactivation can be observed by the monitor using LSM hooks placed at the beginning and the exit of a system call. This way, it becomes possible to track causal dependencies between concurrent system calls within the LSM framework. Laurent Georget has proved (using the Coq proof assistant) that his approach is correct and computes the smallest possible over-approximation, in the sense that for any concurrent execution where multiple system calls are used there exists a linearization of this execution that produces the information flow computed by his algorithm. Laurent Georget has implemented his algorithm in the Linux kernel. This contribution was publish at Software Engineering & Formal Methods (2017) where it was granted the best paper award [11]. Laurent Georget has defended his PhD thesis in September 2017.

Advanced Persistent Threats: Long lived attack campaigns known as Advanced Persistent Threats (APTs) have emerged as a serious security risk. These attack campaigns are customised for their target and performed step by step during months on end. The major difficulty in detecting an APT is keeping track of the different steps logged over months of monitoring and linking them. In [29], we described TerminAPTor, an APT detector which highlights links between the traces left by attackers in the monitored system during the different stages of an attack campaign. TerminAPTor tackles this challenge by resorting to Information Flow Tracking (IFT). TerminAPTor was presented last year and we have pursue our effort in this area. More precisely, we have focus on the evaluation of this solution and thus we face to the lack of public datasets of attacks. We develop Moirai a framework dedicated to attacks scenario sharing [22].

**Characterizing Android Malware**: Android has become the world's most popular mobile operating system, and consequently the most popular target for unscrupulous developers. These developers seek to make money by taking advantage of Android users who customize their devices with various applications, which are the main malware infection vector. Indeed, the most likely way a user executes a repackaged application is by downloading a seemingly harmless application from a store and executing it. Such an application may have been modified by an attacker in order to add malicious pieces of code.

To fight repackaged applications containing malicious code, most official application marketplaces have implemented security analysis tools that try to detect and remove malware. Countermeasures adopted by the attackers to bypass these new controls can be divided into two main approaches: avoiding static analysis and avoiding dynamic analysis. A static analysis of an application consists of analysing its code and its resources without executing it. Conversely, dynamic analysis stands for any kind of analysis that requires executing the application in order to observe its actions.

The Kharon project [30] goes a step further from classical dynamic analysis of malware <sup>0</sup>. Funded by the Labex CominLabs and involving partners of CentraleSupélec, Inria and INSA Centre Val de Loire, this project aims to capture a compact and comprehensive representation of malware. To achieve such a goal we have developed tools to monitor operating systems' information flows induced by the execution of a marked application. We support the idea that the best way to understand malware impact is to observe it in its normal execution environment i.e., a real smartphone. Additionally, the main challenge is to be able to trigger malicious behaviors even when the malware tries to escape dynamic analysis.

In this context, we have developed an original solution whose main purpose is a relavant dynamic analysis of the malicious code. We develop the GroddDroid software, that mainly consists of 'helping the malware to execute'. To reach this goal, GroddDroid relies on a previous static analysis that evidence all the execution paths leading to the malicious code. We compute a global control flow graph (CFG) that exhibits execution paths to reach specific parts of code, even if these paths use callbacks that are handled in the Android framework itself [15]. Finally, GroddDroid slightly modifies the bytecode of the infected application in order to defeat the protection against dynamic analysis and executes the suspicious code in its most favorable execution conditions. Thus, GroddDroid helps to understanding malware's objectives and the consequences on the health of a user's device.

GroddDroid can also be used for classifying applications between goodware and malware. We show in [19] that benign applications have a System Flow Graph (a graph that represent flows at operating system level) that can be anticipated. Malware that perform complex operations such as installing backdoor or launching a Tor client, have a CFG that differ enough to be classified easily.

Our main research direction and challenges in this area are to continue to enhance these technologies in order to reach a sufficient level of software maturity to deploy a permanent platform of malware analysis in the LHS (Laboratory of High Security) and to create new opportunities with industrial partners.

#### 7.1.3. Intrusion Detection in Low-Level Software Components

In order to protect the IDS itself, we have initiated different research activities in the domain of hardware security. Our goal is to use co-design software/hardware approaches against traditional software attacks. In a bilateral research project with HP Inc Research Labs, we investigate how dedicated hardware could be used to monitor the whole software stack (from the firmware to the user-mode applications). In the CominLabs HardBlare project, we study the use of a dedicated co-processor to enforce Information Flow Control (IFC) on the main CPU. Finally, in the context of the PhD thesis of Thomas Letan (ANSSI), we investigate the use of formal methods to evaluate the security guarantees provided by hardware platforms, which combine different CPUs, chipsets and memories.

Highly privileged software, such as firmware, is an attractive target for an attacker. Thus, BIOS vendors use cryptographic signatures to ensure firmware integrity at boot time. Nevertheless, such boot time protection does not prevent an attacker from exploiting vulnerabilities at runtime. To detect such runtime attacks, we proposed an event-based monitoring approach that relies on an isolated co-processor [10]. We instrument the code executed on the main CPU to send information about its behavior to the monitor. In this work, we focus on the detection of attacks targeting the System Management Mode (SMM), a highly privileged x86 execution mode executing firmware code at runtime. We use the control flow of the code as a model of its behavior. We evaluate our approach with two open-source implementations: EDK II and coreboot. We evaluate its ability to detect state-of-the-art attacks and its runtime execution overhead by simulating an x86 system coupled with an ARM Cortex A5 co-processor. The results show that our solution detects intrusions from the state of the art while remaining acceptable in terms of performance overhead in the context of the SMM. This work has been done in collaboration with HP Inc Research Labs, in the context of the PhD of Ronny Chevalier.

Over time, hardware designs have constantly grown in complexity and modern platforms involve multiple interconnected hardware components. During the last decade, several vulnerability disclosures have proven that trust in hardware can be misplaced. The approach we developed with Thomas Letan rely on a formal

<sup>&</sup>lt;sup>0</sup>http://kharon.gforge.inria.fr

definition of Hardware-based Security Enforcement (HSE) mechanisms, a class of security enforcement mechanisms such that a software component relies on the underlying hardware platform to enforce a security policy. We then model a subset of a x86-based hardware platform specifications and we prove the soundness of a realistic HSE mechanism within this model using Coq, a proof assistant system.

The HardBlare project proposes a software/hardware co-design methodology to ensure that security properties are preserved all along the execution of the system but also during files storage. It is based on the Dynamic Information Flow Tracking (DIFT) that generally consists in attaching tags to denote the type of information that are saved or generated within the system. These tags are then propagated when the system evolves and information flow control is performed in order to guarantee the safe execution and storage within the system monitored by security policies. We proposed ARMHEx [20], a practical solution targeting DIFT on ARM-based SoCs (e.g. Xilinx Zynq). Current DIFT implementations suffer from two major drawbacks. First, recovering required information for DIFT is generally based on software instrumentation leading to high time overheads. ARMHEx takes profit of ARM CoreSight debug components and static analysis to drastically reduce instrumentation time overhead (up to 90% compared to existing works). Then, security of the DIFT hardware extension itself is not considered in related works. In this work, we tackle this issue by proposing a solution based on ARM Trustzone. This work has been done in the context of the PhD of Muhammad Abdul Wahab and Mounir Nasr Allah.

#### 7.1.4. Vizualization

When using Intrusion Detection Systems (IDS), the large quantities of alerts generated are difficult to handle by security experts. To help solving this problem, we have proposed VEGAS, an alerts visualization and classification tool that allows primary visions based on their principal component analysis (PCA) representation. Following this, we have studied the context of collaboration between the various security actors. We have then proposed an extension to VEGAS that allows to help the actors to collaborate. We have developped an interface that permits the front-end operator to quickly understand the security events, and group them to organize incidents and send them to dedicated analysts. Conversely, once the incidents have been analysed, the analysts can send information to the front-line operators to help them understanding the futur security events.

We also developed another tool called STARLORD [14] that permits to an administrator the explore in a 3D graph representing the links between the heterogeneous entries in various logs produced either by the system, applications or IDSes. To emphasize the important relations between the lines of logs that can potentially be part of an attack activity, we classify these links in order to present only the part of the graph that is linked to an indicators of compromission.

Our previous research on visualization of security events has lead to two proofs-of-concept (See ELVIS and CORGI softwares). We are currently pursuing business opportunities on this topic. Indeed SplitSec is a soon to be founded startup developing tools to help security experts to better manage and understand security data. Scalable analysis solutions and data visualisations adapted for security are combined into powerful tools for incident response. Until June 2017, Christopher Humphries has been hired by Inria as a technology transfer engineer to build these tools based on promising research prototypes.

#### 7.2. Privacy

#### 7.2.1. Image Encryption

More and more users prefer to share their photos through image-sharing platforms of social networks than using e-mail or personal webpages. Since the provider of the image-sharing platform can clearly know the contents of any published images, the users have to trust the provider to respect their privacy or has to encrypt their images. In the context of the PhD of Kun He, we have proposed an IND-CPA image encryption algorithm that preserve the image format after encryption, and we have shown that our encryption algorithm can be used on several widely used image-sharing platforms such as Flickr, Pinterest, Google+ and Twitter. Kun He has completed her PhD thesis in September 2017 [5].

#### 7.3. Security of Communicating and Distributed Systems

#### 7.3.1. Routing Protocol for Tactical Mobile Ad Hoc Networks

In the context of the PhD thesis of Florian Grandhomme, we propose new secure and efficient algorithms and protocols to provide inter-domain routing in the context of tactical mobile ad hoc network. The proposed protocol has to handle context modification due to the mobility of Mobile Ad hoc NETwork (MANET), that is to say split of a MANET, merge of two or more MANET, and also handle heterogeneity of technology and infrastructure. The solution has to be independent from the underlying intra-domain routing protocol and from the infrastructure: wired or wireless, fixed or mobile. This work is done in cooperation with DGA-MI.

New generation military equipment, soldiers and vehicles, use wireless technology to communicate on the battlefield. During missions, they form a MANET. Since the battlefield includes coalition, each group may communicate with another group, and inter-MANET communication may be established. Inter-MANET (or inter-domain MANET) communication should allow communication, but maintain a control on the exchanged information. Several protocols have been proposed in order to handle inter-domain routing for tactical MANETs. During the thesis we have shown that simulator (NS3) or emulator (CORE) do not handle correctly ad hoc network behavior and then that solution in the state of the art are more complex than needed. Based on this analysis, we propose some preconizations to design Inter-domain protocols for MANET and we propose the ITMAN (Inter Tactical Mobile Ad hoc Network) protocol that allows also to handle simple routing policy (merge, link and deny). We evaluate this new protocol through experimentation and we show that our proposition is quite efficient. On going work on this protocol is the definition and implementation of more subtle routing policy that allow announce filtering of giving prefix for example.

#### 7.3.2. Decentralized Cryptocurrency Systems

Distributed Ledgers (e.g. Bitcoin) occupy currently the first lines of the economical and political media and many speculations are done with respect to their level of coherence and their computability power. Interestingly, there is no consensus on the properties and abstractions that fully capture the behaviour of distributed ledgers. The interest in formalising the behaviour of distributed ledgers is twofold. Firstly, it helps to prove the correctness of the algorithms that implement existing distributed ledgers and explore their limits with respect to an unfriendly environment and target applications. Secondly, it facilitates the identification of the minimal building blocks necessary to implement the distributed ledger in a specific environment. Even though the behaviour of distributed ledgers is similar to abstractions that have been deeply studied for decades in distributed systems no abstraction is sufficiently powerful to capture the distributed ledger behaviour. We have defined the Distributed Ledger Register, a register that mimics the behaviour of one of the most popular distributed ledger, i.e. the Bitcoin ledger. The aim of our work is to provide formal guarantees on the coherent evolution of Bitcoin. We furthermore showed that the Bitcoin blockchain maintenance algorithm verifies the distributed ledger register properties under strict conditions. Moreover, we proved that the Distributed Ledger Register verifies the regularity register specification. It follows that the strongest coherency implemented by Bitcoin is regularity under strong assumptions (i.e. partial synchronous systems and sparse reads). In [7] we proposed a study that contradicts the common belief that Bitcoin implements strong coherency criteria in a totally asynchronous system. To the best of our knowledge, our work is the first one that makes the connection between the distributed ledgers and the classical theory of distributed shared registers.

Double spending and blockchain forks are two main issues that the Bitcoin crypto-system is confronted with. The former refers to an adversary's ability to use the very same coin more than once while the latter reflects the occurrence of transient inconsistencies in the history of the blockchain distributed data structure. We present a new approach to tackle these issues: it consists in adding some local synchronization constraints on Bitcoin's validation operations, and in making these constraints independent from the native blockchain protocol. Synchronization constraints are handled by nodes which are randomly and dynamically chosen in the Bitcoin system. In [13] we show that with such an approach, content of the blockchain is consistent with all validated transactions and blocks which guarantees the absence of both double-spending attacks and blockchain forks.

#### 7.3.3. Large Scale Systems

**Population Protocol**: the computational model of population protocols is a formalism that allows the analysis of properties emerging from simple and pairwise interactions among a very large number of anonymous finitestate agents. Significant work has been done so far to determine which problems are solvable in this model and at which cost in terms of states used by the protocols and time needed to converge. The problem tackled in is the population proportion problem: each agent starts independently from each other in one of two states, say A or B, and the objective is for each agent to determine the proportion of agents that initially started in state A, assuming that each agent only uses a finite set of state, and does not know the number n of agents. In [18], we show that for any  $\delta \in (0, 1)$ , the number of interactions needed per node to converge is  $O(ln(n/\delta))$  with probability at least  $1 - \delta$ . We also prove that each node can determine, with any high probability, the proportion of nodes that initially started in a given state without knowing the number of nodes in the system. This work provides a precise analysis of the convergence bounds, and shows that using the 4-norm is very effective to derive useful bounds.

**Distributed Stream Processing Systems**: shuffle grouping is a technique used by stream processing frameworks to share input load among parallel instances of stateless operators. With shuffle grouping each tuple of a stream can be assigned to any available operator instance, independently from any previous assignment. A common approach to implement shuffle grouping is to adopt a Round-Robin policy, a simple solution that fares well as long as the tuple execution time is almost the same for all the tuples. However, such an assumption rarely holds in real cases where execution time strongly depends on tuple content. As a consequence, parallel stateless operators within stream processing applications may experience unpredictable unbalance that, in the end, causes undesirable increase in tuple completion times. In [25] we propose Online Shuffle Grouping (OSG), a novel approach to shuffle grouping aimed at reducing the overall tuple completion time. OSG estimates the execution time of each tuple, enabling a proactive and online scheduling of input load to the target operator instances. Sketches are used to efficiently store the otherwise large amount of information required to schedule incoming load. We provide a probabilistic analysis and illustrate, through both simulations and a running prototype, its impact on stream processing applications.

The real time analysis of massive data streams is of utmost importance in data intensive applications that need to detect as fast as possible and as efficiently as possible (in terms of computation and memory space) any correlation between its inputs or any deviance from some expected nominal behavior. The IoT infrastructure can be used for monitoring any events or changes in structural conditions that can compromise safety and increase risk. It is thus a recurrent and crucial issue to determine whether huge data streams, received at monitored devices, are correlated or not as it may reveal the presence of attacks. We propose a metric, called codeviation, that allows to evaluate the correlation between distributed massive streams. This metric is inspired from classical metric in statistics and probability theory, and as such enables to understand how observed quantities change together, and in which proportion. In [6], we propose to estimate the codeviation in the data stream model. In this model, functions are estimated on a huge sequence of data items, in an online fashion, and with a very small amount of memory with respect to both the size of the input stream and the values domain from which data items are drawn. We then generalize our approach by presenting a new metric, the Sketch-metric, which allows us to define a distance between updatable summaries of large data streams. An important feature of the Sketch-metric is that, given a measure on the entire initial data streams, the Sketchmetric preserves the axioms of the latter measure on the sketch. We finally conducted extensive experiments on both synthetic traces and real data sets allowing us to validate the robustness and accuracy of our metrics.

## 8. Bilateral Contracts and Grants with Industry

#### 8.1. Bilateral Contracts with Industry

• HP (2013-2019): Embedded Systems Security We aim at researching and prototyping lowlevel intrusion detection mechanisms in embedded system software. This involves mechanisms in continuation of previous work realized by our team as well as investigating new techniques more directly tied to specific HP device architectures. Our main objective is to monitor low-level software (firmware, OS kernels, hypervisors) thanks to a dedicated external co-processor. Ronny Chevalier is doing is PhD in the context of this project. Being under NDA, details about this research program cannot be provided.

#### 8.2. Bilateral Grants with Industry

- Orange Labs: Privacy-preserving location-based services Solenn Brunet has completed her PhD thesis in November 2017 within the context of a CIFRE contract with Orange Labs Caen. Her PhD subject was about privacy-preserving services that are able to provide the service to the user while preserving his privacy. In particular, Solenn Brunet has designed new cryptographic primitives to build anonymous accreditation and she has used these primitives to provide data anonymization mechanisms in the context of e-voting and e-cash.
- DGA: BGP-like Inter Domain routing protocol for tactical mobile ad hoc networks: feasibility, performances and quality of service Florian Grandhomme has completed his PhD thesis in September 2017 in cooperation with DGA-MI. The subject of the PhD was to propose new secure and efficient algorithms and protocols to provide inter-domain routing in the context of tactical mobile ad hoc network. The proposed protocol handles context modification due to the mobility of MANET, that is to say split of a MANET, merge of two or more MANET, and also handles heterogeneity of technology and infrastructure. The solution is independent from the underlying intra-domain routing protocol and from the infrastructure: wired or wireles, fixed or mobile.
- **DGA: Visualization for security events monitoring** Damien Crémilleux has started his PhD thesis in October 2015 in the context of a cooperation with DGA-MI. The subject of the PhD is to define relevant representations to allow front-line security operators to monitors systems from a security perspective. A first proposal was made that led to a tool, VEGAS, that allows to monitor large quantities of alerts in real time and to dispatch these alerts in a relevant way to security analysts.
- DGA: Intrusion Detection in Distributed Applications David Lanoé has started his PhD thesis in October 2016 in the context of a cooperation with DGA-MI. His work is focussing on the construction of behavioral models (during a learning phase) and their use to detect intrusions during an execution of the modelled distributed application.
- Nokia: Risk-aware security policies adaptation in modern communication infrastructures Pernelle Mensah was hired in January 2016 on this CIFRE funding in order to work on unexplored aspects of information security, and in particular response strategies to complex attacks, in the context of cloud computing architectures. The use case proposed by our industrial partner is a multitenant cloud computing platform involving software-defined networking in order to provide further flexibility and responsiveness in architecture management. The topic of the thesis is to adapt and improve the current risk-aware reactive response tools, based on attack graphs and adaptive security policies, to this specific environment, taking into account the heterogeneity of actors, platforms, policies and remediation options.
- Thales: Privacy and Secure Multi-party Computation Aurélien Dupin has started his PhD thesis in January 2016 within the context of a CIFRE contract with Thales. His PhD subject concerns secure multi-party computation. Secure two-party computation provides a way for two parties to compute a function, that depends on the two parties' inputs, while keeping them private. Known since the 1980s, Yao's garbled circuits appear to be a general solution to this problem, in the semi-honest model. Decades of optimizations have made this tool a very practical solution. However, it is well known that a malicious adversary could modify a garbled circuit before submitting it. Many protocols, mostly based on cut-&-choose, have been proposed to secure Yao's garbled circuits in the presence of malicious adversaries. Nevertheless, how much an adversary can modify a circuit and make it still executable have not been studied. In the context of his PhD, Aurélien Dupin is interested by such a question.

- Thales: Combining Attack Specification and Dynamic Learning from traces for correlation rule generation Charles Xosanavongsa has started his PhD thesis in December 2016 in the context of a CIFRE with Thales. His work will focus on the construction of correlation rules. In previous work on correlation rule generation, the usual approach is static. It always relies on the description of the supervised system using a knowledge base of the system. The use of correlation trees is an appealing solution because it allows to have a precise description of the attacks and can handle any kind of IDS. But in practice, the behavior of each IDS is quite difficult to predict, in particular for anomaly based IDS. To manage automatically the correlation rules (and adapt them if necessary), we plan to analyze synthetic traces containing both anomaly based and misused based IDS alerts resulting from an attack.
- Ministry of Defence: Visualisation for the characterization of security events Laetitia Leichtnam has started his PhD thesis in November 2016 in the context of a contract between CentraleSupelec and the French Ministry of Defence. His work consists in presenting events appearing in heterogeneous logs as a dependency graph between the lines of logs. This permits to the administrator to investigate easily the logs to discover the different steps that has performed an attack in the supervised system.
- **ANSSI: Security of Low-level Components** Thomas Letan has started his PhD thesis in the context of a contract between CentraleSupelec and the French National Computer Security Agency (ANSSI). His work consists in using formal methods to specify hardware/software security mechanisms and to verify that they correctly enforce some security policies.

## 9. Partnerships and Cooperations

#### 9.1. Regional Initiatives

- **Region Bretagne ARED Grant :** the PhD of Mourad Leslous on malicious codes in Android applications is supported by a grant from the Région Bretagne.
- Labex COMINLABS contract (2014-2017): "Kharon-Security" http://kharon.gforge.inria.fr

Google Play offers more than 800'000 applications (apps), and this number increases every day. Google play users have performed more than 25 billion app downloads. These applications vary from games to music, video, books, tools, etc. Unfortunately, each of these application is an attack vector on Android. The number of malicious applications (pieces of malware) discovered during the first six months of 2013 exceeds the number of pieces of malware discovered during the 2010 to 2012 period, more than 700 thousand malicious and risky applications were found in the wild. In this context, we propose the Kharon-Security project to stem the progression of Android pieces of malware. We propose to combine static and dynamic monitoring to compute a behavioral signature of Android malware. Behavioral signatures are helpful to understand how malware infect the devices and how they spread information in the Android operating system. Static analysis is essential to understand which particular event or callback triggers malware payload.

In the project we have already developed GroddDroid a tool dedicated to automatic identification and execution of suspicious code. We have also built a dataset of Android malware. In this dataset, all malware are entirely manually reverse and documented. We have also developed an analysis platform. This platform is been deployed at the High Research Laboratory.

• Labex COMINLABS contract (2015-2018): "HardBlare-Security" - http://www.hardblare. cominlabs.ueb.eu/

The general context of the HardBlare project is to address Dynamic Information Flow Tracking (DIFT) that generally consists in attaching marks to denote the type of information that is saved or

generated within the system. These marks are then propagated when the system evolves and information flow control is performed in order to guarantee a safe execution and storage within the system. Existing solutions imply a large overhead induced by the monitoring process. Some attempts rely on a hardware-software approach where DIFT operations are delegated to a coprocessor. Nevertheless, such approaches are based on modified processors. Beyond the fact hardware-assisted DIFT is hardly adopted, existing works do not take care of coprocessor security and multicore/multiprocessor embedded systems.

We plan to implement DIFT mechanisms on boards including a non-modified ARM processor and a FPGA such as those based on the Xilinx Zynq family. The HardBlare project is a multidisciplinary project between CentraleSupélec IETR SCEE research team, CentraleSupélec Inria CIDRE research team and UBS Lab-STICC laboratory. Mounir Nasr Allah is doing his PhD in the context of this project. The main objective of this PhD is to study how hybrid analysis could improve hardware assisted DIFT using static analysis performed at compile-time. Another objective is to manage labels for persistent memory (i.e., files) using a modified OS kernel.

#### Labex COMINLABS contract (2016-2019): "BigClin" - http://www.bigclin.cominlabs.ueb.eu/

Health Big Data (HBD) is more than just a very large amount of data or a large number of data sources. The data collected or produced during the clinical care process can be exploited at different levels and across different domains, especially concerning questions related to clinical and translational research. To leverage these big, heterogeneous, sensitive and multi-domain clinical data, new infrastructures are arising in most of the academic hospitals, which are intended to integrate, reuse and share data for research.

Yet, a well-known challenge for secondary use of HBD is that much of detailed patient information is embedded in narrative text, mostly stored as unstructured data. The lack of efficient Natural Language Processing (NLP) resources dedicated to clinical narratives, especially for French, leads to the development of ad-hoc NLP tools with limited targeted purposes. Moreover, the scalability and real-time issues are rarely taken into account for these possibly costly NLP tools, which make them inappropriate in real-world scenarios. Some other today's challenges when reusing Health data are still not resolved: data quality assessment for research purposes, scalability issues when integrating heterogeneous HBD or patient data privacy and data protection. These barriers are completely interwoven with unstructured data reuse and thus constitute an overall issue which must be addressed globally.

In this project, we plan to develop distributed methods to ensure both the scalability and the online processing of these NLP/IR and data mining techniques; In a second step, we will evaluate the added value of these methods in several real clinical data and on real use-cases, including epidemilology and pharmaco-vigilance, clinical practice assessment and health care quality research, clinical trials.

#### 9.2. National Initiatives

#### 9.2.1. ANR

#### • ANR INFRA Project: SOCIOPLUG (2013-2017) - http://socioplug.univ-nantes.fr/index.php/ SocioPlug\_Project

SocioPlug is a collaborative ANR project involving Inria (ASAP and CIDRE teams), the Nantes University, and LIRIS (INSA Lyon and Université Claude Bernard Lyon). The project emerges from the observation that the features offered by the Web 2.0 or by social media do not come for free. Rather they bring the implicit cost of privacy. Users are more of less consciously selling personal data for services. SocioPlug aims to provide an alternative for this model by proposing a novel architecture for large-scale, user centric applications. Instead of concentrating information of cloud platforms owned by a few economic players, we envision services made possible by cheap low-end plug computers available in every home or workplace. This will make it possible to provide a high

amount of transparency to users, who will be able to decide their own optimal balance between data sharing and privacy.

#### ANR Project: PAMELA (2016-2020) - https://project.inria.fr/pamela/

PAMELA is a collaborative ANR project involving Rennes 1 university (ASAP and CIDRE teams in Rennes), Inria Lille (MAGNET team), LIP6 (MLIA team) and two start-ups, Mediego and Snips. It aims at developing machine learning theories and algorithms in order to learn local and personalized models from data distributed over networked infrastructures. The project seeks to provide first answers to modern information systems built by interconnecting many personal devices holding private user data in the search of personalized suggestions and recommendations. More precisely, we will focus on learning in a collaborative way with the help of neighbors in a network. We aim to lay the first blocks of a scientific foundation for these new types of systems, in effect moving from graphs of data to graphs of data and learned models. CIDRE's contribution in this project involves the design of adversary models and privacy metrics suitable to the privacy-related issues of this distributed learning paradigm.

#### 9.3. International Initiatives

#### 9.3.1. Inria International Partners

#### 9.3.1.1. Informal International Partners

Emmanuelle Anceaume is actively working with Leonardo Querzoni from the University La Sapienza, Italy, on data streams algorithms and engines. Their cooperation gave rise to one publication in Algotel 2017 [25].

Valérie Viet Triem Tong has shortly visited Prof Alexander Pretchner at TU Munchen in june 2017. She has participated to a workshop about Android Malware analysis.

#### 9.4. International Research Visitors

#### 9.4.1. Research Stays Abroad

In the context of the project with HP Inc Labs, Ronny Chevalier and Guillaume Hiet collaborate with the security team of HP Labs in Bristol. They are working more specifically with David Plaquin and Maugan Villatel, who are co-authors of the article published at ASCAC. Ronny Chevalier has spent 3 monts to HP Labs at Bristol.

Mounir Nasr Allah is currently visiting ARM R&D labs at Cambridge for 6 months in the context of the HardBlare project. This visit has been funded by the EIT Digital Doctoral School Program. He is working with Alastair Reid on the use of formal methods to prove that some hardware security mechanisms of ARM embedded processors effectively enforce information flow policies.

Mourad Leslous did an international mobility of three months at the Technical University of Munich, in the team of Professor Alexander Pretschner. This mobility was part of the program of EIT Digital Doctoral School, a European institute that promotes entrepreneurship and innovation among PhD students. During this mobility, he worked on control flow and data flow dependencies in order to detect the malicious code inside Android applications.

## **10.** Dissemination

#### **10.1. Promoting Scientific Activities**

#### 10.1.1. Scientific Events Organisation

#### 10.1.1.1. General Chair, Scientific Chair

Emmanuelle Anceaume co-chair with Maria Potop Butucaru, Christian Cachin and Shlomi Dolev the Workshop on Blockchain Technology and Theory 2017, co-located with DISC 2017.

#### 10.1.1.2. Member of the Organizing Committees

Christophe Bidan served as a member of the organization committee of C&ESAR 2017 (24rd Computers & Electronics Security Applications Rendez-vous), November 2017, Rennes, France.

Frédéric Tronel served as a member of the organization committee of SSTIC 2017 (Symposium sur la sécurité des technologies de l'information et des communications) that took place in Rennes, France from 7th to 9th of June, where it gathered 600 participants.

#### 10.1.2. Scientific Events Selection

#### 10.1.2.1. Member of the Conference Program Committees

Emmanuelle Anceaume served as a member of the following program committees:

- Algotel 2017 (19èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications), May 2017, Quiberon, France.
- ICDCS 2017 (37th IEEE International Conference on Distributed Computing Systems), Atlanta, USA, 2017
- DEBS 2017 (11th ACM International Conference on Distributed and Event-based Systems), Barcelona, Spain 2017
- NCA 2017 (16th International Symposium on Network Computing and Applications), October 2017, Cambridge, MA, USA.
- PECS 2017 (3rd International Conference on Pervasive and Embedded Computing), Porto, Portugal, 2017
- ADSN2017 (6th International Workshop on Assurance in Distributed Systems and Networks, Atlanta, USA, 2017

Christophe Bidan served as a member of the following program committees:

- CRiSIS 2017 (12th International Conference on Risks and Security of Internet and Systems), September 2017, Dinard, France.
- C&ESAR 2017 (24rd Computers & Electronics Security Applications Rendez-vous), November 2017, Rennes, France.

Frédéric Majorczyck served as a member of the program committee of VizSec 2017 (IEEE Symposium on Visualization for Cyber Security), October 2017, Phoeniz, Arizona, USA.

Guillaume Piolle served as a member of the program committee of APVP 2017 (Atelier sur la Protection de la Vie Privée), June 2017, Autrans, France.

Eric Totel served as a member of the program committee of RESSI 2017 (Les Rendez-vous de la recherche et de l'enseignement en sécurité des systèmes d'information).

Frédéric Tronel served as a member of the program committee of SSTIC 2017 (Symposium sur la sécurité des technologies de l'information et des communications) June 2017, Rennes, France.

Gilles guette served as a member of the following program committees:

• ICISSP 2017 (International Conference on Information System Security and Privacy), February 2017, Porto, Portugal.

#### 10.1.2.2. Reviewer

- Gilles Guette IEEE ISNCC 2017 (IEEE International Symposium on Networks, Computers and Communications).
- Michel Hurfin NCA 2017 (16th International Symposium on Network Computing and Applications).
- Jean-François Lalande ICISSP 2017 (3rd International Conference on Information Systems Security and Privacy).

#### 10.1.3. Journal

#### 10.1.3.1. Member of the Editorial Boards

Michel Hurfin belongs to the editorial board of the Springer open access journal of Internet Services and Applications.

#### 10.1.3.2. Reviewer - Reviewing Activities

- Emmanuelle Anceaume Elsevier JPDC (Journal of Parallel and Distributed Computing), Performance Evaluation, IEEE TDSC (Transactions on Dependable and Secure Computing), and IEEE TPDS (Transactions on Parallel and Distributed Systems).
- Michel Hurfin Springer JISA (Journal of Internet Services and Applications) and Springer TOCS (Theory of Computing Systems).
- Jean-François Lalande IARIA IJAS (International Journal on Advances in Security), and Elsevier Computer Communications (International Journal for the Computer and Telecommunications Industry).
- Guillaume Piolle ACM TOIT (Transactions on Internet Technologies).
- Eric Totel IEEE Transactions on Reliability.

#### 10.1.4. Invited Talks

Emmanuelle Anceaume was invited to give

- a keynote at Algotel 2017, May 2017, entitled "Bitcoin and its distributed ledger"
- a keynote at WIFS 2017, December 2017, entitled: "A primer on blockchain technology and the bitcoin cryptocurrency"

#### 10.1.5. Leadership within the Scientific Community

Ludovic Mé serves the Scientific Council of the LIRIMA (Laboratoire International de Recherche en Informatique et Mathématiques Appliquées).

Ludovic Mé chairs the steering Committee of the annual French conference RESSI (Rendez-Vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information). He is a member of the Steering Committee of the annual international conference RAID (International Symposium on Research in Attacks, Intrusions and Defenses).

#### 10.1.6. Research Administration

Emmanuelle Anceaume has participated in various juries (Post-doctoral grants, delegation Inria, PEDR Inria). As a member of the CE Inria, Emmanuelle Anceaume has participated to the hiring committee CR2/CR1 of Rennes and Sophia Antipolis.

Michel Hurfin is the local representative of the "mission jeunes chercheurs" in Rennes. He is a member of the "Commission personnel" and is in charge of the PhD student recruitment campaign of Inria Rennes Bretagne Atlantique. He is a member of the councils of the doctoral school Matisse. He is a member of the advisory board of the doctoral training center of EIT Digital in Rennes.

Ludovic Mé acts as Scientific Officer for the Rennes - Bretagne Atlantic Inria Research Center. As such, he is also a member of the Evaluation Commission and of the Internal Scientific Council of Inria.

Ludovic Mé leads the expert group dedicated to the evaluation of the French laboratories working in the "computing and telecom" domain, relatively to the way they protect their scientific and technical patrimony (PPST French regulation).

Valérie Viet Triem Tong has participated in the scientific evaluation comity *Global Security and Cybersecurity* (CES 39) of the French Research Agency (ANR). Valérie Viet Triem Tong has also participated in the Inria post-doctoral grant.

#### 10.2. Teaching - Supervision - Juries

#### 10.2.1. Certification

The master degree program "Mastère Spécialisé en Cybersécurité" has received the SecNumedu label. This label testifies that this program meets the requirements of a charter that has been jointly established by ANSSI (Agence Nationale de la Sécurité des Systèmes d'Information) and various actors (administrations, companies, ...) of the domain. This label was awarded during the annual Forum International de la Cybersécurité (FIC) in January 2017 at Lille.

#### 10.2.2. Teaching

- Master: Emmanuelle Anceaume, *Research in Computer Science Distributed Algorithms*, 20 hours of lecture, M2; Université Rennes 1, France;
- Licence: Christophe Bidan, *Algorithms and Data Structures*, 36 hours of lecture including 7.5 hours of lectures, L3 first year of the engineering degree, CentraleSupélec, France;
- Licence: Christophe Bidan, *Software Engineering*, 12 hours, L3 first year of the engineer degree, CentraleSupélec, France;
- Licence: Christophe Bidan, *Supervision of student project*, 1 project, L3 first year of the engineer degree, CentraleSupélec, France;
- Master: Christophe Bidan is responsible for the module *Secured information systems*, M2 third year of the engineer degree, CentraleSupélec, France;
- Master: Christophe Bidan, *Applied cryptography*, 6 hours of lecture, M2 master 2 degree, University of Rennes 1, France;
- Master: Christophe Bidan, *Applied cryptography*, 15 hours including 6 hours of lecture, M2 third year of the engineer degree, CentraleSupélec, France;
- Master : Christophe Bidan, *Cryptographic Protocols*, 6 hours of lecture, mastère CS (Cyber Security), CentraleSupélec, France;
- Master: Christophe Bidan, *Information systems*, 4.5 hours, M1 second year of the engineer degree, CentraleSupélec, France;
- Master: Christophe Bidan, *Supervision of student project*, 2 projects, M1 second year of the engineer degree, CentraleSupélec, France;
- Licence: Gilles Guette, *Algorith and Complexity*, 28 hours, L1 Licence, ISTIC/University of Rennes, France;
- Licence: Gilles Guette, *Network Initiation*, 57.5 hours, L3 Licence, ISTIC/University of Rennes, France;
- Licence: Gilles Guette, *Network Initiation*, 41.5 hours, L3 first year of the engineer degree, ESIR, France;
- Master: Gilles Guette, *Network Routing*, 32 hours, M1 second year of the engineer degree, ESIR, France;
- Master: Gilles Guette, *Mobile Network Routing*, 5 hours, M1 second year of the engineer degree, ESIR, France;
- Master: Gilles Guette, *Advanced Network Services*, 10 hours, M1 second year of the engineer degree, ESIR, France;
- Master: Gilles Guette, *Network Project*, 24 hours, M1 second year of the engineer degree, ESIR, France;
- Master: Gilles Guette, Security, 28 hours, M1 second year of the engineer degree, ESIR, France;
- Master: Gilles Guette, *Network and System Security*, 12 hours, M2 third year of the engineer degree, ESIR, France;

- Master: Gilles Guette, *Network Modeling*, 18 hours, M2 third year of the engineer degree, ESIR, France;
- Master: Gilles Guette, *Sensors Network*, 30 hours, M2 Master, ISTIC/University of Rennes, France;
- Master: Gilles Guette, *Supervision of student*, Contrat de professionnalisation, M2 third year of the engineer degree, ESIR, France;
- Master: Gilles Guette, Supervision of student internship, M2 ISTIC/University of Rennes, France;
- Licence: Guillaume Hiet, *Algorithms and Data Structures*, 12.5 hours, L3 first year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Hiet, *Computer security and privacy for the engineer*, 8 hours, M1 second year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Hiet, Buffer overflow vulnerabilities, 16 hours, M2 third year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Hiet, *Buffer overflow vulnerabilities*, 16 hours, M2 Mastère Spécialisé CS (Cyber Security), CentraleSupélec, France;
- Master: Guillaume Hiet, *Pentest*, 19 hours, M2 third year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Hiet, *Pentest*, 3 hours, M2 Mastère Spécialisé CS (Cyber Security), Centrale-Supélec, France;
- Master: Guillaume Hiet, *Introduction to Linux*, 3 hours, M2 Mastère Spécialisé CS (Cyber Security), CentraleSupélec, France;
- Master: Guillaume Hiet, *Java Security*, 4.5 hours, M2 Mastère Spécialisé CS, CentraleSupélec, France;
- Master: Guillaume Hiet, *Linux Security*, 18 hours, M2 Mastère Spécialisé CS, CentraleSupélec, France;
- Master: Guillaume Hiet, *Linux Security*, 7.5 hours, third year of the engineer degree, Centrale-Supélec, France;
- Master: Guillaume Hiet, *LDAP*, 7.5 hours, third year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Hiet, Intrusion Detection, 15 hours, M2 Mastère Spécialisé CS, Centrale-Supélec, France;
- Master: Guillaume Hiet, *Intrusion Detection*, 13.5 hours, M2 third year of the engineer degree, M2 research degree of University of Rennes 1, CentraleSupélec, France;
- Master: Guillaume Hiet, *Security Monitoring*, 3 hours, M2, cycle "Sécurité Numérique", INHESJ, France;
- Master: Guillaume Hiet, *Computer Security*, 31.5 hours, M2, Mastère Spécialisé Architecte des Systèmes d'Information, CentraleSupélec, France;
- Master: Guillaume Hiet, Intrusion Detection, 16 hours, M2, University of Rennes 1, France;
- Master: Guillaume Hiet, *Intrusion Detection*, 10 hours, M2 third year of the engineer degree, ESIR, France;
- Master: Guillaume Hiet, Intrusion Detection, 9 hours, M2, Université of Limoges, France;
- Master: Guillaume Hiet, Firewall, 6 hours, M2, University of Rennes 1, France;
- Master: Guillaume Hiet, Supervision of student project, 4 projects, M1 second year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Hiet, *Supervision of student project*, 2 projects, M2 third year of the engineer degree, CentraleSupélec, France;

- Master: Guillaume Hiet, Supervision of student project, 2 projects, M2 Mastère Spécialisé CS (Cyber Security), CentraleSupélec, France;
- Master: Jean-François Lalande, Legal aspects of information security, 3.5 hours, M2 master CyberSecurity, CentraleSupélec, France;
- Master: Guillaume Hiet, *Android Malware*, 3.5 hours, M2, Mastère Spécialisé CS (Cyber Security), France;
- Master: Jean-François Lalande, Supervision of student project, 2 projects, M1 second year of the engineer degree, CentraleSupélec, France;
- Master: Jean-François Lalande, *Supervision of student project*, 2 projects, M2 third year of the engineer degree, CentraleSupélec, France;
- Master: Jean-François Lalande, Supervision of student project, 1 projects, M2 Mastère Spécialisé CS (Cyber Security), CentraleSupélec, France;
- Licence : Ludovic Mé, *Software Engineering*, 18 hours, L3 first year of the engineer degree, CentraleSupélec, France;
- Licence : Ludovic Mé, *Software Engineering tutorials*, 6 hours, L3 first year of the engineer degree, CentraleSupélec, France;
- Licence : Ludovic Mé, *Software Engineering and Java development*, 18 hours, L3 first year of the engineer degree, CentraleSupélec, France;
- Master : Ludovic Mé, *Information systems tutorials*, 6 hours, M1 second year of the engineer degree, CentraleSupélec, France;
- Master : Ludovic Mé, *Supervision of student project*, 1 project, 38 hours, M1 second year of the engineer degree, CentraleSupélec, France;
- Licence: Guillaume Piolle, *Software engineering*, 1.5 hours, L3 first year of the engineering degree, CentraleSupélec, France;
- Master: Guillaume Piolle, *Modelling, Algorithms and Programming*, 22 hours, M1 second year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Piolle, *Computer security and privacy*, 5 hours, M1 second year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Piolle, *Software project*, 3.5 hours, M1 second year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Piolle, *Relational databases*, 6 hours, M1 second year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Piolle, *Computer networks*, 30 hours, M1 second year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Piolle, *Security Policies*, 4.5 hours, M2 third year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Piolle, *Java programming*, 4.5 hours, M2 third year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Piolle, *Computer networks*, 9 hours, M2 third year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Piolle, *Software engineering*, 12 hours, M2 third year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Piolle, *Network Access Control*, 9 hours, M2 third year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Piolle, *Web development*, 32 hours, M2 third year of the engineer degree, CentraleSupélec, France;

- Master: Guillaume Piolle, *Privacy protection*, 18 hours, M2 third year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Piolle, *Computing project*, 60 hours, M2 third year of the engineer degree, CentraleSupélec, France;
- Master: Guillaume Piolle, *Legal aspects of information security*, 4.5 hours, M2 master CyberSecurity, CentraleSupélec, France;
- Licence : Eric Totel, *Foundations of computer science, data structures and algorithms*, 9 hours, L3 first year of the engineer degree, CentraleSupélec, France;
- Licence : Eric Totel, *Software Modeling*, 15 hours, L3 first year of the engineer degree, Centrale-Supélec, France;
- Master : Eric Totel, *Operating Systems*, 30 hours, M1 second year of the engineer degree, CentraleSupélec, France;
- Master : Eric Totel, *C language*, 24 hours including 6 hours of lecture, M2 master CS (Cyber Security), CentraleSupélec, France;
- Master : Eric Totel, *C language and C++ language*, 12 hours including 6 hours of lecture, M2 third year of the engineer degree, CentraleSupélec, France;
- Master : Eric Totel, *Dependability*, 9 hours including 7.5 hours of lecture, M2 third year of the engineer degree and master research, CentraleSupélec, France;
- Master : Eric Totel, *Dependability*, 3 hours of lecture, M2 third year of the engineer degree (ingénierie des systèmes automatisés), CentraleSupelec, France;
- Master : Eric Totel, *Dependability*, 4.5 hours of lecture, M2 post-graduate training (master Architecture des Réseaux de Communication), CentraleSupélec, France;
- Master : Eric Totel, *Intrusion Detection*, 6 hours of lecture, M2 M2 master CS (Cyber Security), CentraleSupélec, France;
- Master : Eric Totel, *Intrusion Detection*, 9 hours of lecture, M2 master 2 degree, University of Rennes 1, France;
- Master : Eric Totel, *Supervision of student project*, 4 projects, M1 second year of the engineer degree, CentraleSupélec, France;
- Master : Eric Totel, *Supervision of student project*, 1 project, M2 third year of the engineer degree, CentraleSupélec, France;
- Licence: Frédéric Tronel, *Software engineering*, 40 hours, L3 first year of the engineer degree, CentraleSupélec, France;
- Licence: Frédéric Tronel, *Operating Systems*, 12 hours, L3 first year of the engineer degree, CentraleSupélec, France;
- Master: Frédéric Tronel is responsible of the M2 degree in *CyberSecurity* (mastère spécialisé), organized jointly by CentraleSupélec and Institut Mines Télécom (IMT) Atlantique, France;
- Master: Frédéric Tronel, *Operating systems*, 21 hours hours, M2 third year of the engineer degree, CentraleSupélec, France;
- Master: Frédéric Tronel, *Compilers*, 18 hours, M2 third year of the engineer degree, Centrale-Supélec, France;
- Master: Frédéric Tronel, *Automatic reasoning*, 4.5 hours, M2 third year of the engineer degree, CentraleSupélec, France;
- Master: Frédéric Tronel, *Assembly Language*, 6 hours, M2 third year of the engineer degree, CentraleSupélec, France;
- Master: Frédéric Tronel, *Buffer overflow vulnerabilities (theory and practice)*, 20.5 hours, M2 third year of the engineer degree, CentraleSupélec, France;

- Master: Frédéric Tronel, *Firewall*, 15 hours, M2 third year of the engineer degree, CentraleSupélec, France;
- Master: Frédéric Tronel, *Calculability in distributed systems*, 6 hours, M2, jointly with University of Rennes 1 and CentraleSupélec, France;
- Master: Frédéric Tronel, *Computer network*, 8 hours, M2, jointly with University of Rennes 1 and CentraleSupélec, France;
- Licence : Valérie Viet Triem Tong, *Algorithms and Data Structures*, 36 hours of lecture including 7 hours of lectures, L3 first year of the engineering degree, CentraleSupélec, France;
- Licence : Valérie Viet Triem Tong, *Supervision of student project*, 6 projects of 2nd year of the engineer degree, CentraleSupélec, France;
- Master : Valérie Viet Triem Tong, *Games Theory*, 18 hours, M1 second year of the engineering degree, CentraleSupélec, France;
- Master : Valérie Viet Triem Tong, *Formal Methods*, 9 hours, M2 third year of the engineering degree, CentraleSupélec, France;
- Master : Valérie Viet Triem Tong, *Intrusion detection using information flow control*, 9 hours, M2 / third year of the engineering degree, CentraleSupélec, France;
- Master : Valérie Viet Triem Tong, *Programming in Java*, 12 hours, M1 international students (NplusI) second year of the engineering degree, CentraleSupélec, France;
- Master : Valérie Viet Triem Tong, *Small elements of decidability*, 7.5 hours, M2 third year of the engineering degree, CentraleSupélec, France;
- Master : Valérie Viet Triem Tong, *Supervision of student project*, 1 project, mastere CS (Cyber Security), CentraleSupélec, France;
- Master : Valérie Viet Triem Tong, *Supervision of student project*, 8 projects, M1 second year of the engineer degree, CentraleSupélec, France;
- Master : Valérie Viet Triem Tong, *Supervision of student project*, 1 project year of the engineer degre, CentraleSupélec, France;
- Doctorant : Valérie Viet Triem Tong, *Malware analysis by OS information flow tracking*, 2 hours, Summerschool Cyber in Berry, Bourges, France;

#### 10.2.3. Supervision

#### 10.2.3.1. Theses defended in 2017

- PhD: Laurent Georget, *Suivi de flux d'information correct pour les systèmes d'exploitation Linux*, Octobre 2017, supervised by Mathieu Jaume (25% - MdC LIP6), Guillaume Piolle (25%), Frédéric Tronel (25%), and Valérie Viet Triem Tong (25%);
- PhD: Deepak Subramanian, *Multi-level Information Flow Monitoring*, started in January 2013, supervised by Christophe Bidan (20%) and Guillaume Hiet (80%);
- PhD: Antoine Guellier, *Utilisation de la cryptographie homomorphe pour garantir le respect de la vie privée*, started in October 2013, supervised by Christophe Bidan (50%) and Nicolas Prigent (50%);
- PhD: Kun He, *Mise en œuvre de techniques de droit à l'oubli pour les contenus numériques*, started in October 2013, supervised by Christophe Bidan (50%) and Gaetan LeGuelvouit (50% IRT B-Com);
- PhD: Solenn Brunet, *Privacy-preserving location-based services*, started in October 2014, supervised by Christophe Bidan(40%), Sébastien Gambs (30%) and Jacques Traoré (30% - Orange Labs Caen);
- PhD: Florian Grandhomme, *Protocole de routage externe type BGP dans un environnement réseaux tactiques adhoc mobiles : faisabilité et performances*, started in October 2014, supervised by Gilles Guette (50%), Adlen Ksentini (25% Eurecom), and Thierry Plesse (25% DGA MI).

#### 10.2.3.2. Theses in progress

- PhD in progress: Mouna Hkimi, *Détection d'intrusion dans les systèmes distribués*, started in October 2013, supervised by Eric Totel (50%) and Michel Hurfin (50%);
- PhD in progress: Thomas Letan, Contribution à la sécurité des couches basses des systèmes d'information, started in January 2015, supervised by Guillaume Hiet (50%), Pierre Chifflier (25%) ANSSI), and Ludovic Mé (25%);
- PhD in progress: Damien Crémilleux, *Visualisation d'évènements de sécurité pour la supervision*, started in October 2015, supervised by Christophe Bidan (30%), Nicolas Prigent (35%), and Frédéric Majorczyk (35% DGA MI);
- PhD in progress: Mourad Leslous, *Déclenchement automatique de codes jugés suspects dans les applications Android*, started in October 2015, supervised by Thomas Genet (20% Celtique Inria project), Jean François Lalande (40% INSA Centre Val de Loire), and Valérie Viet Triem Tong (40%);
- PhD in progress: Mounir Nasr Allah, *Contrôle de flux d'information par utilisation conjointe d'analyse statique et d'analyse dynamique accélérée matériellement*, started in November 2015, supervised by Guillaume Hiet (75%) and Ludovic Mé (25%);
- PhD in progress: Pernelle Mensah, *Adaptation de la Politique de Sécurité guidée par l'Évaluation du Risque dans les Infrastructures de Communication modernes*, started in January 2016, supervised by Eric Totel (25%), Guillaume Piolle (25%), Christine Morin (25% Myriad Inria project), and Samuel Dubus (25% Nokia);
- PhD in progress: David Lanoë, *Détection d'intrusion dans les applications distribuées : l'approche comportementale comme alternative à la corrélation d'alertes*, started in october 2016, supervised by Michel Hurfin (50%) and Eric Totel (50%);
- PhD in progress: Aurélien Trulla, *Caractérisation de malware Android par suivi de flux d'information et nouvelles techniques d'évasion*, started in October 2016, supervised by Jean Louis Lanet (25% Tamis Inria project) and Valérie Viet Triem Tong (75%);
- PhD in progress : Ronny Chevalier , "Enhanced computer platform security through an intrusiondetection approach", started in November 2016, supervised by Guillaume Hiet (50%), Boris Balacheff (25% - HP), and Ludovic Mé (25%);
- PhD in progress: Laetitia Leichtnam, *Visualisation pour la caractérisation d'événements de sécurité*, started in october 2016, supervised by Eric Totel (40%), Nicolas Prigent (30%) and Ludovic Mé (30%);
- PhD in progress : Charles Xosanavongsa, *Combining Attack Specification and Dynamic Learning from traces for correlation rule generation*, started in december 2016, supervised by Eric Totel (50%) and Ludovic Mé (50%);
- PhD in progress : Pierre Graux, *Security of Hybrid Mobile Applications*, started in october 2017, supervised by Valérie Viet Triem Tong (50%) and Jean-Francçois Lalande (50%);
- PhD in progress : Vasile Cazacu, *Calcul distribué pour la fouille de données cliniques*, started February 2017, supervised by Emmanuelle Anceaume (50%) and Marc Cuggia (50%)
- PhD in progress : Aurélien Dupin, *Secure multi-partie computations*, started February 2016, supervised by Christophe Bidan(40%), David Pointchavalm (30% ENS) and Renaud Dubois (30% Thales).

#### 10.2.3.3. Supervision of external PhD candidates

• LL. D. (Doctor of Laws) in progress: Gustav Malis, *Droit à l'effacement des données mises à disposition par les personnes elles-mêmes*, started in March 2014, supervised by Annie Blandin (80% - IODE) and Guillaume Piolle (20%);

- PhD in progress: Oualid Koucham, *Détection d'intrusions pour les systèmes de contrôle industriels*, started in January 2015, supervised by Stéphane Mocanu (50% Gipsa-lab), Guillaume Hiet (25%), and Jean-Marc Thiriet (25% Gipsa-lab);
- PhD in progress : Yves Mocquard, *Population protocols*, started in september 2015, supervised by Bruno Sericola (Dyonisos Inria project) and Emmanuelle Anceaume;.

#### 10.2.4. Juries

Ludovic Mé was a member of the PhD committee for the following PhD and HDR thesis:

• Pierre Laperdrix, Browser Fingerprinting: Diversity to Augment Authentication and Build Clientside Countermesures, INSA of Rennes, 03/10/2017 (President of the Jury);

Ludovic Mé has reported the following PhD thesis:

• Pierre Parrend, *Emergent Industrial Ecosystems*, University of Strasbourg, 12/12/2017.

Christophe Bidan was a member of the PhD committee for the following PhD thesis:

- Jean Aimé Maxa, Architecture de communication sécurisée d'une flotte de drones, Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier), 28/06/2017;
- Eric Asselin, Système de détection d'intrusion adapté au système de communication aéronautique ACARS, Institut National Polytechnique de Toulouse (INP Toulouse), 28/06/2017 (reviewer);

Jean-François Lalande has reported the following PhD thesis:

• Julien Hatin, Evaluation de la confiance dans un processus d'authentification, 24/11/2017.

Eric Totel was a member of the PhD committee for the following PhD thesis:

- Giannakou Anna, *Self-Adaptable Security Monitoring for IaaS Cloud Environments*, 06/07/2017 (President of the Jury).
- Yacine Hebbal, Semantic monitoring mechanisms dedicated to security monitoring in IaaS cloud, 18/09/2017 (President of the Jury).

Valérie Viet Triem Tong has reported the following PhD thesis:

 Franck de Goer de Herve, Rétro-ingénierie de programmes binaires en une exécution - une analyse dynamique légère basée au niveau des fonctions, 10/20/2017.

#### **10.3.** Popularization

Valérie Viet Triem Tong has participated to the scientific television show *l'Esprit Sorcier* recorded at *Musée* des Sciences et de l'Industrie during the Fête de la Science. She has also participated to the scientific promotion movie about *High Security Laboratory* recorded at Nancy.

Damien Crémilleux has participated to the event "Ma thèse en 180s" and the "RCC challenge: my thesis 3.0" for the popularization of his work's thesis on security visualization.

## 11. Bibliography

#### **Publications of the year**

#### **Doctoral Dissertations and Habilitation Theses**

- [1] S. BRUNET. *Design of Anonymous Accreditation and Data Anonymization Mechanisms*, Université de Rennes 1 [UR1], November 2017, https://hal-centralesupelec.archives-ouvertes.fr/tel-01655322.
- [2] L. GEORGET. Correct information flow tracking for Linux operating systems, Université Rennes 1, September 2017, https://hal.inria.fr/tel-01657148.

- [3] F. GRANDHOMME.Inter-domain routing studies (BGP-like) for tactical mobile ad hoc networks : feasibility and performances, Universite de Rennes 1, November 2017, https://hal.archives-ouvertes.fr/tel-01654631.
- [4] A. GUELLIER. *Strongly Private Communications in a Homogeneous Network*, Centrale Supélec, May 2017, https://hal-centralesupelec.archives-ouvertes.fr/tel-01644172.
- [5] K. HE.Content Privacy and Access Control on Image-Sharing Platforms, CentraleSupélec, September 2017, https://hal-centralesupelec.archives-ouvertes.fr/tel-01636207.

#### **Articles in International Peer-Reviewed Journal**

[6] E. ANCEAUME, Y. BUSNEL.Lightweight Metric Computation for Distributed Massive Data Streams, in "Transactions on Large-Scale Data- and Knowledge-Centered Systems", April 2017, vol. 10430, n<sup>o</sup> 33, p. 1–39 [DOI: 10.1007/978-3-662-55696-2\_1], https://hal.archives-ouvertes.fr/hal-01634353.

#### **International Conferences with Proceedings**

- [7] E. ANCEAUME, R. LUDINARD, M. POTOP-BUTUCARU, F. TRONEL.*Bitcoin a Distributed Shared Register*, in "SSS 2017 - 19th International Symposium on Stabilization, Safety, and Security of Distributed Systems", Boston, MA, United States, Lecture Notes in Computer Science, Springer, November 2017, vol. 10616, p. 456-468 [DOI: 10.1007/978-3-319-69084-1\_34], https://hal.archives-ouvertes.fr/hal-01522360.
- [8] M. ASSAF, D. A. NAUMANN, J. SIGNOLES, E. TOTEL, F. TRONEL. Hypercollecting Semantics and its Application to Static Analysis of Information Flow, in "POPL 2017 - ACM Symposium on Principles of Programming Languages", Paris, France, ACM SIGPLAN Notices - POPL '17, ACM, January 2017, vol. 52, n<sup>o</sup> 1, p. 874-887 [DOI: 10.1145/3009837.3009889], https://hal.inria.fr/hal-01618360.
- [9] I. BOUREANU, D. GERAULT, P. LAFOURCADE, C. ONETE.*Breaking and fixing the HB+DB protocol*, in "Wisec 2017 - Conference on Security and Privacy in Wireless and Mobile Networks", Boston, United States, July 2017, p. 241 - 246 [DOI: 10.1145/3098243.3098263], https://hal.archives-ouvertes.fr/hal-01588562.
- [10] R. CHEVALIER, M. VILLATEL, D. PLAQUIN, G. HIET.Co-processor-based Behavior Monitoring: Application to the Detection of Attacks Against the System Management Mode, in "ACSAC 2017 33rd Annual Computer Security Applications Conference", Orlando, United States, Proceedings of the 33rd Annual Computer Security Applications Conference, ACM, December 2017, vol. 2017, p. 399-411 [DOI: 10.1145/3134600.3134622], https://hal.inria.fr/hal-01634566.

#### [11] Best Paper

L. GEORGET, M. JAUME, G. PIOLLE, F. TRONEL, V. VIET TRIEM TONG.*Information Flow Tracking for Linux Handling Concurrent System Calls and Shared Memory*, in "15th International Conference on Software Engineering and Formal Methods (SEFM 2017)", Trento, Italy, A. CIMATTI, M. SIRJANI (editors), LNCS, Springer International Publishing, September 2017, p. 1-16 [*DOI* : 10.1007/978-3-319-66197-1\_1], http://hal.upmc.fr/hal-01535949.

[12] L. GEORGET, M. JAUME, G. PIOLLE, F. TRONEL, V. VIET TRIEM TONG. Verifying the Reliability of Operating System-Level Information Flow Control Systems in Linux, in "5th International FME Workshop on Formal Methods in Software Engineering", Buenos Aires, Argentina, IEEE Press, May 2017, p. 10-16 [DOI: 10.1109/FORMALISE.2017.1], https://hal.inria.fr/hal-01535862.

- [13] T. LAJOIE-MAZENC, R. LUDINARD, E. ANCEAUME.*Handling Bitcoin Conflicts Through a Glimpse of Structure*, in "Proceedings of the 32nd ACM SIGAPP Symposium On Applied Computing", Marrakesh, Morocco, Proceedings of the 32nd ACM SIGAPP Symposium On Applied Computing, April 2017 [DOI: 10.1145/3019612.3019657], https://hal.archives-ouvertes.fr/hal-01634368.
- [14] L. LEICHTNAM, E. TOTEL, N. PRIGENT, L. MÉ.STARLORD: Linked Security Data Exploration in a 3D Graph, in "VizSec - IEEE Symposium on Visualization for Cyber Security", Phoenix, United States, October 2017, p. 1 - 4 [DOI: 10.1109/VIZSEC.2017.8062203], https://hal.inria.fr/hal-01619234.
- [15] M. LESLOUS, V. VIET TRIEM TONG, J.-F. LALANDE, T. GENET.GPFinder: Tracking the Invisible in Android Malware, in "12th International Conference on Malicious and Unwanted Software", Fajardo, Puerto Rico, IEEE Conputer Society, October 2017, https://hal-centralesupelec.archives-ouvertes.fr/hal-01584989.
- [16] P. MENSAH, S. DUBUS, W. KANOUN, C. MORIN, G. PIOLLE, E. TOTEL. Connectivity Extraction in Cloud Infrastructures, in "CNSM 2017 - 13th International Conference on Network and Service Management", Tokyo, Japan, November 2017, p. 1-5, https://hal.inria.fr/hal-01593346.
- [17] P. MENSAH, S. DUBUS, W. KANOUN, C. MORIN, G. PIOLLE, E. TOTEL. Connectivity Graph Reconstruction for Networking Cloud Infrastructures, in "NCA 2017 - 16th IEEE International Symposium on Network Computing and Applications", Cambridge, MA, United States, October 2017, https://hal.inria.fr/hal-01612988.
- [18] Y. MOCQUARD, B. SERICOLA, E. ANCEAUME. Probabilistic Analysis of Counting Protocols in Largescale Asynchronous and Anonymous Systems, in "Proceedings of the 16th IEEE International Conference on Network Computing and Applications", Boston, United States, October 2017, https://hal.archives-ouvertes.fr/ hal-01580275.
- [19] V. VIET TRIEM TONG, A. TRULLA, M. LESLOUS, J.-F. LALANDE.*Information flows at OS level unmask sophisticated Android malware*, in "14th International Conference on Security and Cryptography", Madrid, Spain, SciTePress, July 2017, vol. 6, p. 578-585 [DOI: 10.5220/0006476705780585], https:// hal-centralesupelec.archives-ouvertes.fr/hal-01535678.
- [20] M. A. WAHAB, P. COTRET, M. N. ALLAH, G. HIET, V. LAPOTRE, G. GOGNIAT.ARMHEx: A hardware extension for DIFT on ARM-based SoCs, in "Field Programmable Logic (FPL)", Ghent, Belgium, September 2017, https://hal.archives-ouvertes.fr/hal-01558473.
- [21] M. A. WAHAB, P. COTRET, M. NASR ALLAH, G. HIET, V. LAPOTRE, G. GOGNIAT.ARMHEx: a framework for efficient DIFT in real-world SoCs, in "Field Programmable Logic (FPL)", Ghent, Belgium, September 2017, https://hal.archives-ouvertes.fr/hal-01558475.

#### **National Conferences with Proceeding**

- [22] G. BROGI, V. VIET TRIEM TONG. Sharing and replaying attack scenarios with Moirai, in "RESSI 2017: Rendez-vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information", Autrans, France, May 2017, https://hal.archives-ouvertes.fr/hal-01533275.
- [23] G. FOURNIER, P. AUDREN DE KERDREL, P. COTRET, V. VIET TRIEM TONG.DroneJack: Kiss your drones goodbye!, in "SSTIC 2017 - Symposium sur la sécurité des technologies de l'information et des communications", Rennes, France, June 2017, p. 1-8, https://hal.inria.fr/hal-01635125.
- [24] L. GEORGET, M. JAUME, G. PIOLLE, F. TRONEL, V. VIET TRIEM TONG. Suivi de flux d'information correct sous Linux, in "16èmes journées AFADL (Approches formelles dans l'assistance au développement de logiciels)", Montpellier, France, A. IDANI, N. KOSMATOV (editors), June 2017, p. 19-26, http://hal.upmc. fr/hal-01535937.
- [25] N. RIVETTI, E. ANCEAUME, Y. BUSNEL, L. QUERZONI, B. SERICOLA. Ordonnancement dynamique pour un équilibrage de charge quasi-optimal dans les systèmes de traitement de flux, in "ALGOTEL 2017 - 19èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Quiberon, France, May 2017, https://hal.archives-ouvertes.fr/hal-01519432.
- [26] M. A. –. WAHAB, P. COTRET, M. –. NASR ALLAH, G. –. HIET, V. LAPOTRE, G. –. GOGNIAT.ARMHEx: embedded security through hardware-enhanced information flow tracking, in "RESSI 2017 : Rendez-vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information", Grenoble (Autrans), France, May 2017, https://hal.archives-ouvertes.fr/hal-01558155.

#### **Other Publications**

- [27] E. ANCEAUME, Y. BUSNEL. Byzantine-tolerant Uniform Node Sampling Service, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01634363.
- [28] Y. MOCQUARD, B. SERICOLA, E. ANCEAUME. *Probabilistic Analysis of Rumor Spreading Time*, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01652777.

#### **References in notes**

- [29] G. BROGI, V. VIET TRIEM TONG. TerminAPTor: Highlighting Advanced Persistent Threats through Information Flow Tracking, in "8th IFIP International Conference on New Technologies, Mobility and Security", Larnaca, Cyprus, November 2016, https://hal.inria.fr/hal-01417612.
- [30] N. KISS, J.-F. LALANDE, M. LESLOUS, V. VIET TRIEM TONG. *Kharon dataset: Android malware under a microscope*, in "The Learning from Authoritative Security Experiment Results (LASER) workshop", San Jose, United States, The USENIX Association, May 2016, https://hal-centralesupelec.archives-ouvertes.fr/hal-01311917.

# **Project-Team DIONYSOS**

# Dependability Interoperability and perfOrmance aNalYsiS Of networkS

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: IMT Atlantique Bretagne-Pays de la Loire

Université Rennes 1

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Networks and Telecommunications

## **Table of contents**

1.	Personnel	185
2.	Overall Objectives	186
3.	Research Program	187
	3.1. Introduction	187
	3.2. Quality of Service and Quality of Experience	187
	3.3. Stochastic modeling	188
4.	Application Domains	. 189
	4.1. Networking	189
_	4.2. Stochastic modeling	189
5.	Highlights of the Year	189
6.	New Software and Platforms	. 189
	6.1. IPvb lest loolkit	189
	6.2. Passive lest lool	190
	6.3. I 3DevKit	190
	6.4. ttproto	190
	6.5. COAP lesting lool	190
	6.6. ioppytest	191
	6.7. AdaComp	191
-	0.8. DININ-WIINKINL	191
7.	New Results   7.1   Desformence Evaluation	192
	7.1. Performance Evaluation 7.2. Distributed deep learning on edge devices	192
	7.2. Distributed deep learning on edge-devices	194
	7.5. Network Economics	194
	7.4. Windess Networks	190
	7.5. Which is Networks	100
	7.7 Future networks and architectures	200
8	Bilateral Contracts and Grants with Industry	200
0.	8.1 Cifre contract on Device-Assisted Distributed Machine-Learning on Many Cores	203
	8.2 Cifre contract on Throughput Prediction in Mobile Networks	203
	8.3. Cifre contract on Mobile SDN architecture	203
	8.4. Cifre contract on Personalization for Cognitive Autonomic Networks in 5G	203
	8.5. Bilateral Contract with Industry: ALSTOM-Inria Common Lab	203
	8.6. Bilateral Contract with Industry: ADR Nokia Bell Labs	204
9.	Partnerships and Cooperations	204
	9.1. National Initiatives	204
	9.1.1. ANR	204
	9.1.2. IPL (Inria Project Lab) BetterNet	204
	9.2. European Initiatives	204
	9.2.1. Eurostars Camion Project	204
	9.2.2. Collaborations in European Programs	205
	9.2.3. Collaborations with Major European Organizations	205
	9.3. International Initiatives	205
	9.3.1. Inria International Partners	205
	9.3.2. Participation in Other International Programs	205
	9.3.2.1. International Initiatives	205
	9.3.2.2. International Initiatives	206
	9.4. International Research Visitors	206
10.	Dissemination	. 206

10.1. Promoting Scientific Activities	206
10.1.1. Scientific Events Organisation	206
10.1.2. Scientific Events Selection	207
10.1.2.1. Chair of Conference Program Committees	207
10.1.2.2. Member of the Conference Program Committees	207
10.1.2.3. Reviewer	208
10.1.3. Journal	208
10.1.3.1. Member of the Editorial Boards	208
10.1.3.2. Reviewer - Reviewing Activities	209
10.1.4. Invited Talks	209
10.1.5. Leadership within the Scientific Community	210
10.1.6. Scientific Expertise	210
10.1.7. Research Administration	210
10.2. Teaching - Supervision - Juries	210
10.2.1. Teaching	210
10.2.2. Supervision	211
10.2.3. Juries	212
10.3. Popularization	213
11. Bibliography	213

#### **Project-Team DIONYSOS**

Creation of the Project-Team: 2009 January 01

#### **Keywords:**

#### **Computer Science and Digital Science:**

- A1.1.6. Cloud
- A1.1.7. Peer to peer
- A1.1.13. Virtualization
- A1.2.2. Supervision
- A1.2.3. Routing
- A1.2.4. QoS, performance evaluation
- A1.2.5. Internet of things
- A1.3. Distributed Systems
- A3.4.1. Supervised learning
- A3.4.2. Unsupervised learning
- A3.4.3. Reinforcement learning
- A3.4.6. Neural networks
- A3.4.8. Deep learning
- A6.1.1. Continuous Modeling (PDE, ODE)
- A6.2.2. Numerical probability
- A6.2.3. Probabilistic methods
- A6.2.6. Optimization
- A8.1. Discrete mathematics, combinatorics
- A8.11. Game Theory

#### **Other Research Topics and Application Domains:**

B1.2.1. - Understanding and simulation of the brain and the nervous system

- B2.2. Physiology and diseases
- B6.2.1. Wired technologies
- B6.2.2. Radio technology
- B6.2.4. Optic technology
- B6.3.2. Network protocols
- B6.4. Internet of things

## **1. Personnel**

#### **Research Scientists**

Gerardo Rubino [Team leader, Inria, Senior Researcher, HDR] Pierre L'Ecuyer [Univ. of Montreal, Canada, and Inria Advanced Research Position] Raymond Marie [Univ de Rennes I (Emeritus), Senior Researcher, HDR] Bruno Sericola [Inria, Senior Researcher, HDR] Bruno Tuffin [Inria, Senior Researcher, HDR]

#### **Faculty Members**

Yann Busnel [IMT Atlantique, Professor, HDR]

Yassine Hadjadj-Aoul [Univ de Rennes I, Associate Professor] Sofiene Jelassi [Univ de Rennes I, Associate Professor] Patrick Maille [IMT Atlantique, Professor, HDR] César Viho [Univ de Rennes I, Professor, HDR]

#### **Technical Staff**

Pantelis Frangoudis [Univ de Rennes I, until Jan 2017] Quang Pham Tran Anh [Univ de Rennes I, from Jul 2017 until Sep 2017] Federico Sismondi [Univ de Rennes I]

#### **PhD Students**

Imad Alawe [IRT B-com] Hamza Ben Ammar [Univ de Rennes I] Corentin Hardy [Technicolor] Ali Hodroj [Univ de Rennes I] Nicolas Jara Carvallo [Ingeniería Civil Telemática UTFSM, until Sep 2017] Farouk Messaoudi [Institut de recherche technologique B-com, until Sep 2017] Yves Mocquard [Univ de Rennes I] Mohamed Rahali [from Oct 2017] Ajit Rai [ALSTOM] Illyyne Saffar [Univ de Rennes I] Imane Taibi [Inria, from Nov 2017] Louiza Yala [Univ de Rennes I] Yue Li [Inria, until Feb 2017] Jean-Michel Sanner [Orange Lab]

#### Visiting Scientists

Nawel Attia [Technische Universität München, Apr 2017] Jorge Graneri [Universidad de la República de Uruguay, from Oct 2017 until Nov 2017] Gustavo Guerberoff [Universidad de la República de Uruguay, from Oct 2017 until Nov 2017] Leslie Murray [University of Rosario, Argentina, Feb 2017] Jonathan Paolo Olavarria Milla [UTFSM, Chile, until Mar 2017] Claudio Risso [Universidad de la República de Uruguay, from Oct 2017 until Nov 2017]

#### Administrative Assistant

Fabienne Cuyollaa [Inria]

## 2. Overall Objectives

#### 2.1. Overall objectives

The main objectives of the project are the identification, the conception and the selection of the most appropriate network architectures for a communication service, as well as the development of computing and mathematical tools for the fulfillment of these tasks. These objectives lead to two types of complementary research fields: the systems' qualitative aspects (e.g. protocol testing and design) and the quantitative aspects which are essential to the correct dimensioning of these architectures and the associated services (performance, dependability, Quality of Service (QoS), Quality of Experience (QoE) and performability); our activities lie essentially in the latter.

The Dionysos group works on different problems related to the design and the analysis of communication services. Such services require functionality specifications, decisions about where and how they must be deployed in a system, and the dimensioning of their different components. The interests of the project concern not only particular classes of systems but also methodological aspects.

Concerning the communication systems themselves, we focus on IP networks, at different levels. Concerning the types of networks considered, we mainly work in the wireless area, in particular on sensor networks, on Content Delivery Networks for our work around measuring the perceived quality, the main component of QoE, and on some aspects of optical networks. We also work on the assessment of interoperability between specific network components, which is essential to ensure that they interact correctly before they get deployed in a real environment. Our team contributes in providing solutions (methods, algorithms and tools) which help in obtaining efficient interoperability test suites for new generation networks. From the application point of view, we also have activities in network economics methodologies, a critical multi-disciplinary area for telecommunications providers, with many defying open problems for the near future.

For most of previous mentioned problems, our work concern their quantitative aspects. The quantitative aspects we are interested in are QoE, performance, dependability, performability, QoS, vulnerability, etc. We develop techniques for the evaluation of these different aspects of the considered systems through *models* and through *measurement techniques*. In particular, we develop techniques to measure in an automatic way the quality of a video or audio communication *as perceived by the final user*. The methods we work with range from discrete event simulation and Monte Carlo procedures to analytical techniques, and include numerical algorithms as well. Our main mathematical tools are stochastic processes in general and queuing models and Markov chains in particular, optimization techniques, graph theory, combinatorics, etc.

## **3. Research Program**

#### 3.1. Introduction

The scientific foundations of our work are those of network design and network analysis. Specifically, this concerns the principles of packet switching and in particular of IP networks (protocol design, protocol testing, routing, scheduling techniques), and the mathematical and algorithmic aspects of the associated problems, on which our methods and tools are based.

These foundations are described in the following paragraphs. We begin by a subsection dedicated to Quality of Service (QoS) and Quality of Experience (QoE), since they can be seen as unifying concepts in our activities. Then we briefly describe the specific sub-area of model evaluation and about the particular multidisciplinary domain of network economics.

#### 3.2. Quality of Service and Quality of Experience

Since it is difficult to develop as many communication solutions as possible applications, the scientific and technological communities aim towards providing general *services* allowing to give to each application or user a set of properties nowadays called "Quality of Service" (QoS), a terminology lacking a precise definition. This QoS concept takes different forms according to the type of communication service and the aspects which matter for a given application: for performance it comes through specific metrics (delays, jitter, throughput, etc.), for dependability it also comes through appropriate metrics: reliability, availability, or vulnerability, in the case for instance of WAN (Wide Area Network) topologies, etc.

QoS is at the heart of our research activities: We look for methods to obtain specific "levels" of QoS and for techniques to evaluate the associated metrics. Our ultimate goal is to provide tools (mathematical tools and/or algorithms, under appropriate software "containers" or not) allowing users and/or applications to attain specific levels of QoS, or to improve the provided QoS, if we think of a particular system, with an optimal use of the resources available. Obtaining a good QoS level is a very general objective. It leads to many different areas, depending on the systems, applications and specific goals being considered. Our team works on several of these areas. We also investigate the impact of network QoS on multimedia payloads to reduce the impact of congestion.

Some important aspects of the behavior of modern communication systems have subjective components: the quality of a video stream or an audio signal, *as perceived by the user*, is related to some of the previous mentioned parameters (packet loss, delays, ...) but in an extremely complex way. We are interested in analyzing these types of flows from this user-oriented point of view. We focus on the *user perceived quality*, in short, PQ, the main component of what is nowadays called Quality of Experience (in short, QoE), to underline the fact that, in this case, we want to center the analysis on the user. In this context, we have a global project called PSQA, which stands for Pseudo-Subjective Quality Assessment, and which refers to a technology we have developed allowing to automatically measure this PQ.

Another special case to which we devote research efforts in the team is the analysis of qualitative properties related to interoperability assessment. This refers to the act of determining if end-to-end functionality between at least two communicating systems is as required by the base standards for those systems. Conformance is the act of determining to what extent a single component conforms to the individual requirements of the standard it is based on. Our purpose is to provide such a formal framework (methods, algorithms and tools) for interoperability assessment, in order to help in obtaining efficient interoperability test suites for new generation networks, mainly around IPv6-related protocols. The interoperability test suites generation is based on specifications (standards and/or RFCs) of network components and protocols to be tested.

#### 3.3. Stochastic modeling

The scientific foundations of our modeling activities are composed of stochastic processes theory and, in particular, Markov processes, queuing theory, stochastic graphs theory, etc. The objectives are either to develop numerical solutions, or analytical ones, or possibly discrete event simulation or Monte Carlo (and Quasi-Monte Carlo) techniques. We are always interested in model evaluation techniques for dependability and performability analysis, both in static (network reliability) and dynamic contexts (depending on the fact that time plays an explicit role in the analysis or not). We look at systems from the classical so-called *call level*, leading to standard models (for instance, queues or networks of queues) and also at the *burst level*, leading to *fluid models*.

In recent years, our work on the design of the topologies of WANs led us to explore optimization techniques, in particular in the case of very large optimization problems, usually formulated in terms of graphs. The associated methods we are interested in are composed of simulated annealing, genetic algorithms, TABU search, etc. For the time being, we have obtained our best results with GRASP techniques.

Network pricing is a good example of a multi-disciplinary research activity half-way between applied mathematics, economy and networking, centered on stochastic modeling issues. Indeed, the Internet is facing a tremendous increase of its traffic volume. As a consequence, real users complain that large data transfers take too long, without any possibility to improve this by themselves (by paying more, for instance). A possible solution to cope with congestion is to increase the link capacities; however, many authors consider that this is not a viable solution as the network must respond to an increasing demand (and experience has shown that demand of bandwidth has always been ahead of supply), especially now that the Internet is becoming a commercial network. Furthermore, incentives for a fair utilization between customers are not included in the current Internet. For these reasons, it has been suggested that the current flat-rate fees, where customers pay a subscription and obtain an unlimited usage, should be replaced by usage-based fees. Besides, the future Internet will carry heterogeneous flows such as video, voice, email, web, file transfers and remote login among others. Each of these applications requires a different level of QoS: for example, video needs very small delays and packet losses, voice requires small delays but can afford some packet losses, email can afford delay (within a given bound) while file transfer needs a good average throughput and remote login requires small round-trip times. Some pricing incentives should exist so that each user does not always choose the best QoS for her application and so that the final result is a fair utilization of the bandwidth. On the other hand, we need to be aware of the trade-off between engineering efficiency and economic efficiency; for example, traffic measurements can help in improving the management of the network but is a costly option. These are some of the various aspects often present in the pricing problems we address in our work. More recently, we have switched to the more general field of network economics, dealing with the economic behavior of users, service providers and content providers, as well as their relations.

## 4. Application Domains

#### 4.1. Networking

Our global research effort concerns networking problems, both from the analysis point of view, and around network design issues. Specifically, this means the IP technology in general, with focus on specific types of networks seen at different levels: wireless systems, optical infrastructures, peer-to-peer architectures, Software Defined Networks, Content Delivery Networks, Content-Centric Networks, clouds.

A specific aspect of network applications and/or services based on video or voice content, is our PSQA technology, able to measure the Perceptual Quality automatically and in real time. PSQA provides a MOS value as close as it makes sense to the value obtained from subjective testing sessions. The technology has been tested in many environments, including one way communications as, for instance, in video streaming, and bi-directional communications as in IP telephony, UDP- or TCP-based systems, etc. It has already served in many collaborative projects as the measuring tool used.

#### 4.2. Stochastic modeling

Many of the techniques developed at Dionysos are related to the analysis of complex systems in general, not only in telecommunications. For instance, our Monte Carlo methods for analyzing rare events have been used by different industrial partners, some of them in networking but recently also by companies building transportation systems. We develop methods in different areas: numerical analysis of stochastic models, bound computations in the same area, Discrete Event Simulation, or, as just mentioned, rare event analysis.

## 5. Highlights of the Year

#### 5.1. Awards

BEST PAPER AWARD : [43] IEEE NCA 2017 - 16th IEEE International Symposium on Network Computing and Applications. C. HARDY, E. LE MERRER, B. SERICOLA.

## 6. New Software and Platforms

#### 6.1. IPv6 Test Toolkit

FUNCTIONAL DESCRIPTION: These test suites are developed using the TTCN-3 environment.

The packages contains the full Abstract Test Suites written in TTCN-3 and the source files for building the codecs and adapters with the help of T3DevKit.

- Participants: Annie Floch, Anthony Baire, Ariel Sabiguero, Bruno Deniaud, César Viho and Frédéric Roudaut
- Contact: César Viho

#### **6.2.** Passive Test Tool

- Participants: Anthony Baire and César Viho
- Contact: Anthony Baire

### 6.3. T3DevKit

KEYWORDS: IPv6 - Conformance testing - TTCN-3

SCIENTIFIC DESCRIPTION: We have built a toolkit for easing executing tests written in the standardized TTCN-3 test specification language. This toolkit is made of a C++ library together with a highly customizable CoDec generator that allows fast development of external components (that are required to execute a test suite) such as CoDec (for message Coding/Decoding), System and Platform Adapters. It also provides a framework for representing and manipulating TTCN-3 events so as to ease the production of test reports. The toolkit addresses issues that are not yet covered by ETSI standards while being fully compatible with the existing standard interfaces: TRI (Test Runtime Interfaces) and TCI (Test Control Interfaces), it has been tested with four TTCN-3 environments (IBM, Elvior, Danet and Go4IT) and on three different platforms (Linux, Windows and Cygwin).

FUNCTIONAL DESCRIPTION: T3DevKit is a free open source toolkit to ease the development of test suites in the TTCN-3 environment. It provides:

a CoDec generator (t3cdgen) that automates the development process of the CoDec needed for coding TTCN-3 values into physically transmittable messages and decoding incoming messages a library (t3devlib) that provides an object oriented framework to manipulate TTCN-3 entities (values, ports, timers, external functions...) an implementation of the TRI and TCI standard interfaces default implementations for the system adapter (SA), platform adapter (PA), test management (TM), test logging (TL) and component handling (CH) modules default codecs build scripts for the generation of executable test suites, these are tool-independent and facilitate the distribution of test suite sources

- Participants: Annie Floch, Anthony Baire, Ariel Sabiguero, César Viho and Frédéric Roudaut
- Contact: Federico Sismondi

#### 6.4. ttproto

#### Testing Tool Prototype

KEYWORDS: Interoperability - Conformance testing - TTCN-3

FUNCTIONAL DESCRIPTION: ttproto is an experimental tool for implementing testing tools, for conformance and interoperability testing.

It was first implemented to explore new features and concepts for the TTCN-3 standard, but we also used it to implement a passive interoperability test suite we provided for the CoAP interoperability event held in Paris in March 2012.

This tool is implemented in python3 and its design was influenced mainly by TTCN-3 (abstract model, templates, snapshots, behaviour trees, communication ports, logging) and by Scapy (syntax, flexibility, customisability)

Its purpose is to facilitate rapid prototyping rather than experimentations (rather than production use). We choosed to maximise its modularity and readability rather than performances and real-time considerations.

Now you should have a look at the Features page: https://www.irisa.fr/tipi/wiki/doku.php/ testing\_tool\_prototype:features

- Contact: Federico Sismondi
- URL: https://www.irisa.fr/tipi/wiki/doku.php/testing\_tool\_prototype

#### 6.5. CoAP Testing Tool

KEYWORDS: Test - Interoperability - Conformance testing - Plugtests

FUNCTIONAL DESCRIPTION: The software helps developers of the CoAP protocol assessing if their implementations (either CoAP clients or CoAP servers) are conformant to protocol specifications, and interoperable with other implementations. It encompasses:

- Coordination of CoAP interoperability tests
- Analysis of CoAP traces & issuing verdicts
- Automation of open source CoAP implementations for based reference interop testing
- Authors: Federico Sismondi and César Viho
- Contact: Federico Sismondi

#### 6.6. ioppytest

#### Interoperability testing

KEYWORDS: Interoperability - Conformance testing - CoAP - 6LoWPAN - OneM2M

FUNCTIONAL DESCRIPTION: The software is a framework for developing interoperability tests. The interoperability tests help developers of network protocol assessing if their implementations are conformant to protocol specifications, and interoperable with other implementations.

The software already integrates interoperability tests for CoAP, OneM2M and 6LoWPAN The framework provides the following features to the users:

- Coordination of the interoperability tests (enabling remote testing )
- VPN-like connectivity between users' implementations (enabling remote testing )
- Analysis of exchanged network traces & issuing verdicts
- Automation of open source implementations for based reference interop testing

This framework is the evolution of the CoAP Testing Tool (https://bil.inria.fr/fr/software/view/2937/tab)

- Contact: Federico Sismondi
- URL: https://gitlab.f-interop.eu/f-interop-contributors/ioppytest

#### 6.7. AdaComp

#### Participants: Corentin Hardy, Bruno Sericola

Our recent works, in collaboration with Technicolor, on deep learning and distributed learning led us to study a kind of data parallelism called the Parameter Server model. This model consists in sharing the learning of a deep neural network between many devices (called the workers) via a centralized Parameter Server (PS). We deployed a platform which allow us to experiment different state-of-the-art algorithms based on the PS model. The platform is composed of a unique powerful machine where many Linux containers (LXC) are running. Each LXC executes a Tensorflow session and can be a worker or a PS. The first experimentations were used to validate the correct functioning of the platform, to better understand its limitations and to determine what can be measured in an unbiased way. Others experimentations helped us to understand the role of different parameters of the overall model, mainly those related to the distribution on user-devices, and their impact on the learning (accuracy of the model, number of iterations to learn the model). During these experimentations, we noted that the main bottleneck is the ingress traffic of PS during the learning phase. To reduce this ingress traffic, we chose to compress the messages sent by the workers to the PS. We proposed in [43] a method to reduce up to 2 orders of magnitude this ingress traffic, keeping a good accuracy on the learned model. This new method, called AdaComp, is available in github (https://github.com/Hardy-c/AdaComp).

#### 6.8. DNN-withRNL

Participants: Corentin Hardy, Gerardo Rubino, Bruno Sericola

The extension of the AdaComp method, presented in 6.7, to Random Neural Networks started with the introduction of Random Neural Layers, see [65]. Concerning the associated software, see https://github.com/Hardy-c/DNN-with-RNL.

## 7. New Results

#### 7.1. Performance Evaluation

Participants: Yann Busnel, Yves Mocquard, Bruno Sericola, Gerardo Rubino

Correlation estimation between distributed massive streams. The real time analysis of massive data streams is of utmost importance in data intensive applications that need to detect as fast as possible and as efficiently as possible (in terms of computation and memory space) any correlation between its inputs or any deviance from some expected nominal behavior. The IoT infrastructure can be used for monitoring any events or changes in structural conditions that can compromise safety and increase risk. It is thus a recurrent and crucial issue to determine whether huge data streams, received at monitored devices, are correlated or not as it may reveal the presence of attacks. In [14] we propose a metric, called *Codeviation*, that allows to evaluate the correlation between distributed massive streams. This metric is inspired from classical material in statistics and probability theory, and as such enables to understand how observed quantities change together, and in which proportion. We then propose to estimate the codeviation in the data stream model. In this model, functions are estimated on a huge sequence of data items, in an online fashion, and with a very small amount of memory with respect to both the size of the input stream and the domain from which data items are drawn. We then generalize our approach by presenting a new metric, the *Sketch-*\$\$\preservec\$ metric, which allows us to define a distance between updatable summaries of large data streams. An important feature of the Sketch-radiate metric is that, given a measure on the entire initial data streams, the *Sketch-*\$\sim metric preserves the axioms of the latter measure on the sketch. We also conducted extensive experiments on both synthetic traces and real data sets allowing us to validate the robustness and accuracy of our metrics.

**Stream processing systems.** Stream processing systems are today gaining momentum as tools to perform analytics on continuous data streams. Their ability to produce analysis results with sub-second latencies, coupled with their scalability, makes them the preferred choice for many big data companies.

A stream processing application is commonly modeled as a direct acyclic graph where data operators, represented by nodes, are interconnected by streams of tuples containing data to be analyzed, the directed edges (the arcs). Scalability is usually attained at the deployment phase where each data operator can be parallelized using multiple instances, each of which will handle a subset of the tuples conveyed by the operators' ingoing stream. Balancing the load among the instances of a parallel operator is important as it yields to better resource utilization and thus larger throughputs and reduced tuple processing latencies.

*Shuffle grouping* is a technique used by stream processing frameworks to share input load among parallel instances of stateless operators. With shuffle grouping each tuple of a stream can be assigned to any available operator instance, independently from any previous assignment. A common approach to implement shuffle grouping is to adopt a Round-Robin policy, a simple solution that fares well as long as the tuple execution time is almost the same for all the tuples. However, such an assumption rarely holds in real cases where execution time strongly depends on tuple content. As a consequence, parallel stateless operators within stream processing applications may experience unpredictable unbalance that, in the end, causes undesirable increase in tuple completion times. In [61] we propose Online Shuffle Grouping (OSG), a novel approach to shuffle grouping a proactive and online scheduling of input load to the target operator instances. Sketches are used to efficiently store the otherwise large amount of information required to schedule incoming load. We provide a probabilistic analysis and illustrate, through both simulations and a running prototype, its impact on stream processing applications.

*Grand Challenge*. Since 2011, the ACM International Conference on Distributed Event-based Systems (DEBS) launched the Grand Challenge series to increase the focus on these systems as well as provide common benchmarks to evaluate and compare them. The ACM DEBS 2017 Grand Challenge focused on (soft) real-time anomaly detection in manufacturing equipment. To handle continuous monitoring, each machine is fitted with a vast array of sensors, either digital or analog. These sensors provide periodic measurements, which are sent to a monitoring base station. The latter receives then a large collection of observations. Analyzing in an efficient and accurate way, this very-high-rate – and potentially massive – stream of events is the core of the Grand Challenge. Although, the analysis of a massive amount of sensor reading requires an on-line analytics pipeline that deals with linked-data, clustering as well as a Markov model training and querying. The FlinkMan system [62] proposes a solution to the 2017 Grand Challenge, making use of a publicly available streaming engine and thus offering a generic solution that is not specially tailored for this or for another challenge. We offer an efficient solution that maximally utilizes available cores, balances the load among the cores, and avoids to the extent possible tasks such as garbage collection that are only indirectly related to the task at hand.

*Health big data processing.* Sharing and exploiting efficiently Health Big Data (HBD) lead to tackle great challenges: data protection and governance taking into account legal, ethical and deontological aspects which enables a trust, transparent and win-to-win relationship between researchers, citizen and data providers. Lack of interoperability: data are compartmentalized and are so syntactically and semantically heterogeneous. Variable data quality with a great impact on data management and statistical analysis. The objective of the INSHARE project [41] is to explore, through an experimental proof of concept, how recent technologies could overcome such issues. It aims at demonstrating the feasibility and the added value of an IT platform based on CDW, dedicated to collaborative HBD sharing for medical research.

The consortium includes 6 data providers: 2 academic hospitals, the SNIIRAM (the French national reimbursement database) and 3 national or regional registries. The platform is designed following a three steps approach: (1) to analyze use cases, needs and requirements, (2) to define data sharing governance and secure access to the platform, (3) to define the platform specifications. Three use cases (healthcare trajectory analysis, epidemiological registry enrichment, signal detection) were analyzed to design the platform corresponding to five studies and using eleven data sources. The governance was derived from the SCANNER model and adapted to data sharing. As a result, the platform architecture integrates the following tools and services: data repository and hosting, semantic integration services, data processing, aggregate computing, data quality and integrity monitoring, id linking, multi-source query builder, visualization and data export services, data governance, study management service and security including data watermarking.

**Throughput prediction in cellular networks.** Downlink data rates can vary significantly in cellular networks, with a potentially non-negligible effect on the user experience. Content providers address this problem by using different representations (*e.g.*, picture resolution, video resolution and rate) of the same content and by switching among these based on measurements collected during the connection. If it were possible to know the achievable data rate before the connection establishment, content providers could choose the most appropriate representation from the very beginning. We have conducted a measurement campaign involving 60 users connected to a production network in France, to determine whether it is possible to predict the achievable data rate using measurements collected, before establishing the connection to the content provider, on the operator's network and on the mobile node. We show that it is indeed possible to exploit these measurements to predict, with a reasonable accuracy, the achievable data rate [53].

**Population protocol model.** We consider in [50] a large system populated by n anonymous nodes that communicate through asynchronous and pairwise interactions. The aim of these interactions is, for each node, to converge toward a global property of the system that depends on the initial state of the nodes. We focus on both the counting and proportion problems. We show that for any  $\delta \in (0, 1)$ , the number of interactions needed per node to converge is  $O(\ln(n/\delta))$  with probability at least  $1 - \delta$ . We also prove that each node can determine, with any high probability, the proportion of nodes that initially started in a given state without knowing the number of nodes in the system. This work provides a precise analysis of the convergence bounds, and shows that using the 4-norm is very effective to derive useful bounds.

The context of [71] is the well studied dissemination of information in large scale distributed networks through pairwise interactions. This problem, originally called *rumor mongering*, and then *rumor spreading* has mainly been investigated in the synchronous model, which relies on the assumption that all the nodes of the network act in synchrony, that is, at each round of the protocol, each node is allowed to contact a random neighbor. In this paper, we drop this assumption under the argument that it is not realistic in large scale systems. We thus consider the asynchronous variant, where, at random times, nodes successively interact by pairs exchanging their information on the rumor. In a previous paper, we performed a study of the total number of interactions needed for all the nodes of the network to discover the rumor. While most of the existing results involve huge constants that do not allow us to compare different protocols, we provided a thorough analysis of the distribution of this total number of interactions together with its asymptotic behavior. In this paper we extend this discrete-time analysis by solving a conjecture proposed previously and we consider the continuous-time case, where a Poisson process is associated with each node to determine the instants at which interactions occur. The rumor spreading time is thus more realistic since it is the time needed for all the nodes of the network to discover the rumor. Once again, as most of the existing results involve huge constants, we provide a tight bound and equivalent of the complementary distribution of the rumor spreading time. We also give the exact asymptotic behavior of the complementary distribution of the rumor spreading time around its expected value when the number of nodes tends to infinity.

**Transient analysis.** Last, in two keynotes ([35] and [34]), we described part of our previous analytical results concerning the transient behavior of well-structured Markov processes, mainly on performance models (queueing systems), and we presented recent new results that extend those initial findings. The heart of the novelties lie on an extension of the concept of duality proposed by Anderson in [73] that we call pseudo-dual. The dual of a stochastic process needs strong monotonicity conditions to exist. Our proposed pseudo-dual always exist, and is directly defined on a linear system of differential equations with constant coefficients, that can be, in particular, the system of Chapman-Kolmogorov equations corresponding to a Markov process, but not necessarily. This allows, for instance, to prove the validity of closed-forms expressions of the transient distribution of a Markov process in cases where the dual doesn't exist. The keynote [35] was presented to a public oriented toward differential equations and dynamical systems; [34] has a more modeling flavour. A paper is under preparation with the technical details.

#### 7.2. Distributed deep learning on edge-devices

Participants: Corentin Hardy, Gerardo Rubino, Bruno Sericola

A large portion of data mining and analytic services use modern machine learning techniques, such as deep learning. The state-of-the-art results related to deep learning come at the price of an intensive use of computing resources. The leading frameworks (e.g., TensorFlow) are executed on GPUs or on high-end servers in data centers. On the other end, there is a proliferation of personal devices with possibly free CPU cycles; this can enable services to run in users' homes, embedding machine learning operations. In [66] and [43], we ask the following question: Is distributed deep learning computation on WAN connected devices feasible, in spite of the traffic caused by learning tasks? We show that such a setup rises some important challenges, most notably the ingress traffic that the servers hosting the up-to-date model have to sustain. In order to reduce this stress, we propose AdaComp, a novel algorithm for compressing worker updates to the model on the server. Applicable to stochastic gradient descent based approaches, it combines efficient gradient selection and learning rate modulation. We experiment and measure the impact of compression, device heterogeneity and reliability on the accuracy of learned models, with an emulator platform that embeds TensorFlow into Linux containers. We report a reduction of the total amount of data sent by workers to the server by two order of magnitude (e.g., 191-fold reduction for a convolutional network on the MNIST dataset), when compared to a standard asynchronous stochastic gradient descent, while preserving model accuracy. The extension of the AdaComp algorithm to Random Neural Networks started with the introduction of Random Neural Layers, see [65].

#### 7.3. Network Economics

Participants: Bruno Tuffin, Patrick Maillé, Pierre L'Ecuyer

The general field of network economics, analyzing the relationships between all acts of the digital economy, has been an important subject for years in the team. The whole problem of network economics, from theory to practice, describing all issues and challenges, is described in our book [7].

**Roaming.** In October 2015, the European parliament has decided to forbid roaming charges among EU mobile phone users, starting June 2017, as a first step toward the unification of the European digital market. We have investigated the consequences of such a measure from an economic perspective. In [47], we analyze the effect of the willingness-to-pay heterogeneity among users (also due to wealth heterogeneity), and the fact that the roaming behavior is positively correlated with wealth. Our analysis suggests that imposing free roaming degrades the revenues of the operator but can also deter some users from subscribing; hence we conclude that such (apparently beneficial) regulatory decisions must be taken with care. In [47], we particularly focus on the strategies on transit payments between ISPs in different countries. We highlight that scrutiny is also required since, depending on parameters, consumer surplus or subscription penetration are not necessarily maximized if free roaming is enforced.

**Network neutrality.** Most of our activity has been devoted to the vivid network neutrality debate, going beyond the traditional for or against neutrality, and trying to tackle it from different angles.

Network neutrality has been a very sensitive topic of discussion all over the world. In the keynote talk [59], we first introduce the elements of the debate and how the problem can be modeled and analyzed through game theory. With an Internet ecosystem much more complex now than the simple delivery chain Content-ISP-User, we highlight, in a second step, how neutrality principles can be bypassed in various ways without violating the rules currently evoked in the debate, for example via Content Delivery Networks (CDNs), or via search engines which can affect the visibility and accessibility of content. We describe some other grey zones requiring to be dealt with and spend some time on discussing the (potential) implications for clouds.

The impact of CDNs on the debate has been detailed in [18]. Content Delivery Networks (CDN) have become key telecommunication actors. They contribute to improve significantly the quality of services delivering content to end users. However, their impact on the ecosystem raises concerns about their fairness, and therefore the question of their inclusion in the neutrality debates becomes relevant. We analyze the impact of a revenue-maximizing CDN on some other major actors, namely, the end-users, the network operators and the content providers, at comparing the outcome with that of a fair behavior, and at providing tools to investigate whether some regulation should be introduced. We present a mathematical model and show that there exists a unique optimal revenue-maximizing policy for a CDN actor, in terms of dimensioning and allocation of its storage capacity, and depending on parameters such as prices for service/transport/storage. Numerical experiments are then performed with both synthetic data and real traces obtained from a major Video-on-Demand provider. In addition, using the real traces, we compare the revenue-based policy with policies based on several fairness criteria.

Network neutrality is often advocated by content providers, stressing that side payments to Internet Service Providers would hinder innovation. However, we also observe some content providers actually paying those fees. In [24], we intend to explain such behaviors through economic modeling, illustrating how side payments can be a way for an incumbent content provider to prevent new competitors from entering the market. We investigate the conditions under which the incumbent can benefit from such a barrier-to-entry, and the consequences of that strategic behavior on the other actors: content providers, users, and the Internet Service Provider. We also describe how the Nash bargaining solution concept can be used to determine the side payment.

Similarly, major content/service providers are publishing grades they give to ISPs about the quality of delivery of their content. The goal is to inform customers about the "best" ISPs. But this could be an incentive for, or even a pressure on, ISPs to differentiate service and provide a better quality to those big content providers in order to be more attractive. Instead of the traditional vision of ISPs pressing content providers, we face here the opposite situation, still possibly at the expense of small content providers though. We design in [48] a model describing the various actors and their strategies, analyzes it using non-cooperative game theory tools, and quantifies the impact of those advertised grades with respect to the situation where no grade is published. We illustrate that a non-neutral behavior, differentiating traffic, is not leading to a desirable situation.

**Sponsored data**. With wireless sponsored data, a third party, content or service provider, can pay for some of your data traffic so that it is not counted in your plan's monthly cap. This type of behavior is currently under scrutiny, with telecommunication regulators wondering if it could be applied to prevent competitors from entering the market, and what the impact on all telecommunication actors can be. To answer those questions, we design and analyze in [69] a model where a Content Provider (CP) can choose the proportion of data to sponsor and a level of advertisement to get a return on investment, and several Internet Service Providers (ISPs) in competition. We distinguish three scenarios: no sponsoring, the same sponsoring to all users, and a different sponsoring depending on the ISP you have subscribed to. This last possibility may particularly be considered an infringement of the network neutrality principle. We see that sponsoring can be beneficial to users and ISPs, especially with identical sponsoring. We also discuss the impact of zero-rating where an ISP offers free data to a CP to attract more customers, of and vertical integration where a CP and an ISP are the same company.

**Online platforms and search engines**. The search neutrality debate is about whether search engines should or should not be allowed to uprank certain results among the organic content matching a query. This debate is related to that of network neutrality, which focuses on whether all bytes being transmitted through the Internet should be treated equally. In a previous paper, we had formulated a model that formalizes this question and characterized an optimal ranking policy for a search engine. The model relies on the trade-off between short-term revenues, captured by the benefits of highly-paying results, and long-term revenues which can increase by providing users with more relevant results to minimize churn. In [21], we apply that model to investigate the relations between search neutrality and innovation. We illustrate through a simple setting and computer simulations that a revenue-maximizing search engine may indeed deter innovation at the content level. Our simple setting obviously simplifies reality, but this has the advantage of providing better insights on how optimization by some actors impacts other actors.

**Sponsored auctions.** Advertisement in dedicated webpage spaces or in search engines sponsored slots is usually sold using auctions, with a payment rule that is either per impression or per click. But advertisers can be both sensitive to being viewed (brand awareness effect) and being clicked (conversion into sales). In [23], we generalize the auction mechanism by including both pricing components: the advertisers are charged when their ad is displayed, and pay an additional price if the ad is clicked. Applying the results for Vickrey-Clarke-Groves (VCG) auctions, we show how to compute payments to ensure incentive compatibility from advertisers as well as maximize the total value extracted from the advertisement slot(s). We provide tight upper bounds for the loss of efficiency due to applying only pay-per-click (or pay-per-view) pricing instead of our scheme. Those bounds depend on the joint distribution of advertisement visibility and population likelihood to click on ads, and can help identify situations where our mechanism yields significant improvements. We also describe how the commonly used generalized second price (GSP) auction can be extended to this context.

#### 7.4. Monte Carlo

#### Participants: Bruno Tuffin, Gerardo Rubino, Pierre L'Ecuyer

We maintain a research activity in different areas related to dependability, performability and vulnerability analysis of communication systems, using both the Monte Carlo and the Quasi-Monte Carlo approaches to evaluate the relevant metrics. Monte Carlo (and Quasi-Monte Carlo) methods often represent the only tool able to solve complex problems of these types. We have published an introduction to Monte Carlo methods on Insterstices, including animations https://interstices.info/jcms/int\_69164/la-simulation-de-monte-carlo.

**Rare event simulation**. The mean time to failure (MTTF) of a stochastic system is often estimated by simulation. One natural estimator, which we call the direct estimator, simply averages independent and identically distributed copies of simulated times to failure. When the system is regenerative, an alternative approach is based on a ratio representation of the MTTF. The purpose of [42] is to compare the two estimators. We first analyze them in the setting of crude simulation (i.e., no importance sampling), showing that they are actually asymptotically identical in a rare-event context. The two crude estimators are inefficient in different but closely related ways: the direct estimator requires a large computational time because times to failure often include many transitions, whereas the ratio estimator entails estimating a rare-event probability. We then

discuss the two approaches when employing importance sampling; for highly reliable Markovian systems, we show that using a ratio estimator is advised.

Another problem studied in [40] is the estimation of the tail of the distribution of the sum of correlated log-normal random variables. While a number of theoretically efficient estimators have been proposed for this setting, using a few numerical examples we illustrate that these published proposals may not always be useful in practical simulations. As a remedy to this defect, we propose a new estimator and we demonstrate that, not only is our novel estimator theoretically efficient, but, more importantly, its practical performance is significantly better than that of its competitors.

**Random variable generation.** Random number generators were invented before there were symbols for writing numbers, and long before mechanical and electronic computers. All major civilizations through the ages found the urge to make random selections, for various reasons. Today, random number generators, particularly on computers, are an important (although often hidden) ingredient in human activity. In the invited paper [32], we give a historical account on the design, implementation, and testing of uniform random number generators used for simulation.

We study in [68] the lattice structure of random number generators of the specific MIXMAX family, a class of matrix linear congruential generators that produce a vector of random numbers at each step. These generators were initially proposed and justified as close approximations to certain ergodic dynamical systems having the Kolmogorov K-mixing property, which implies a chaotic (fast-mixing) behavior. But for a K-mixing system, the matrix must have irrational entries, whereas for the MIXMAX it has only integer entries. As a result, the MIXMAX has a lattice structure just like linear congruential and multiple recursive generators. We study this lattice structure for vectors of successive and non-successive output values in various dimensions. We show in particular that for coordinates at specific lags not too far apart, in three dimensions, all the nonzero points lie in only two hyperplanes. This is reminiscent of the behavior of lagged-Fibonacci and AWC/SWB generators. And even if we skip the output coordinates involved in this bad structure, other highly structured projections often remain, depending on the choice of parameters.

**Quasi-Monte Carlo (QMC).** In [5], which appeared in 2017, we survey basic ideas and results on randomized quasi-Monte Carlo (RQMC) methods, discuss their practical aspects, and give numerical illustrations. RQMC can improve accuracy compared with standard Monte Carlo (MC) when estimating an integral interpreted as a mathematical expectation. RQMC estimators are unbiased and their variance converges at a faster rate (under certain conditions) than MC estimators, as a function of the sample size. Variants of RQMC also work for the simulation of Markov chains, for function approximation and optimization, for solving partial differential equations, etc. In this introductory survey, we look at how RQMC point sets and sequences are constructed, how we measure their uniformity, why they can work for high-dimensional integrals, and how can they work when simulating Markov chains over a large number of steps.

**General presentations.** Finally, in two general presentations, we described state-of-the-art technologies available to deal with rare events by means of Monte Carlo techniques, including several methods produced inside Dionysos. In the tutorial [33], we gave an overview of the field, with a focus on dependability analysis applications. The keynote [36] described specific procedures taken from our monograph [72], that were adapted to the needs of the micro-simulation community.

#### 7.5. Wireless Networks

**Participants:** Yue Li, Imad Alawe, Quang Pham, Patrick Maillé, Yassine Hadjadj-Aoul, César Viho, Gerardo Rubino

**Mobile wireless networks' improvements.** Software Defined Networking (SDN) is one of the key enablers for evolving mobile network architecture towards 5G. SDN involves the separation of control and data plane functions, which leads, in the context of 5G, to consider the separation of the control and data plane functions of the different gateways of the Evolved Packet Core (EPC), namely Serving and Packet data Gateways (S and P-GW). Indeed, the envisioned solutions propose to separate the S/P-GW into two entities: the S/P-GW-C, which integrates the control plane functions and the S/P-GW-U that handles the User Equipment (UE)

data plane traffic. There are two major approaches to create and update user plane forwarding rules for such a partition: (i) considering an SDN controller for the S/P-GW-C (SDNEPC) or (ii) using a direct specific interface to control the S/P-GW-U (enhancedEPC). In [38], we evaluate, using a testbed, those two visions against the classical virtual EPC (vEPC), where all the elements of the EPC are virtualized. Besides evaluating the capacity of the vEPC to manage and scale to UE requests, we compare the performances of the solutions in terms of the time needed to create the user data plane. The obtained results allow drawing several remarks, which may help to dimension the vEPC's components as well as to improve the S/P-GW-U management procedure.

One of the requirements of 5G is to support a massive number of connected devices, considering many usecases such as IoT and massive Machine Type Communication (MTC). While this represents an interesting opportunity for operators to grow their business, it will need new mechanisms to scale and manage the envisioned high number of devices and their generated traffic. Particularity, the signaling traffic, which will overload the 5G core Network Function (NF) in charge of authentication and mobility, namely Access and Mobility Management Function (AMF). The objective of [37] is to provide an algorithm based on Control Theory allowing: (i) to equilibrate the load on the AMF instances in order to maintain an optimal response time with limited computing latency; (ii) to scale out or in the AMF instance (using NFV techniques) depending on the network load to save energy and avoid wasting resources. Obtained results indicate the superiority of our algorithm in ensuring fair load balancing while scaling dynamically with the traffic load. In [64] we are going further by using new advances on machine learning, and more specifically Recurrent Neural Networks (RNN), to predict accurately the arrival traffic pattern of devices. The main objective of the proposed approach is to early react to congestion by pro-actively scaling the AMF VNF in a way to absorb such congestion while respecting the traffic constraints.

**Energy consumption improvements.** Recently in cellular networks, the focus has been moved to seeking ways to increase the energy efficiency by better adapting to the existing users behaviors. In [17], we are going a step further in studying a new type of disruptive service by trying to answer the question "What are the potential energy efficiency gains if some of the users are willing to tolerate delays?". We present an analytical model of the energy usage of LTE base stations, which provides lower bounds of the possible energy gains under a decentralized, noncooperative setup. The model is analyzed in six different scenarios (such as micromacro cell interaction and coverage redundancy) for varying traffic and user-tolerable delays. We show that it is possible to reduce the power consumption by up to 30%.

**Computation offloading in mobile network.** Mobile edge computing (MEC) emerges as a promising paradigm that extends the cloud computing to the edge of pervasive radio access networks, in near vicinity to mobile users, reducing drastically the latency of end-to-end access to computing resources. Moreover, MEC enables the access to up-to-date information on users' network quality via the radio network information service (RNIS) application programming interface (API), allowing to build novel applications tailored to users' context. In [25] and [49], we present a novel framework for offloading computation tasks, from a user device to a server hosted in the mobile edge (ME) with highest CPU availability. Besides taking advantage of the proximity of the MEC server, the main innovation of the proposed solution is to rely on the RNIS API to drive the user equipment (UE) decision to offload or not computing tasks for a given application. The contributions are twofold. First, we propose the design of an application hosted in the ME, which estimates the current value of the round trip time (RTT) between the UE and the ME, according to radio quality indicators available through RNIS API, and provides it to the UE. Second, we present a novel computation algorithm which, based on the estimated RTT coupled with other parameters (e.g., energy consumption), decide when to offload UE's applications computing tasks to the MEC server. The effectiveness of the proposed framework is demonstrated via testbed experiments featuring a face recognition application.

**Services improvement in wireless heterogeneous networks.** With the rapid growth of HTTP-based Adaptive Streaming (HAS) multimedia video services on the Internet, improving the Quality of Experience (QoE) of video delivery will be highly requested in wireless heterogeneous networks. Various access technologies such as 3G/LTE and Wi-Fi with overlapping coverage is the main characteristic of network heterogeneity. Since contemporary mobile devices are usually equipped with multiple radio interfaces, mobile users are enabled to

utilize multiple access links simultaneously for additional capacity or reliability. However, network and video quality selection can have notable impact on the QoE of DASH clients facing the video service's requirements, the wireless channel profiles and the costs of the different links. In this context, the emerging Multi-access Edge Computing (MEC) standard gives new opportunities to improve DASH performance, by moving IT and cloud computing capabilities down to the edge of the mobile network. In [45], we propose a MEC-assisted architecture for improving the performance of DASH-based streaming, a standard implementation of a HAS framework in wireless heterogeneous networks. With the proposed algorithm running as a MEC service, the overall QoE and fairness of DASH clients are improved in a real time manner in case of network congestion.

**QoE aware routing in wireless networks.** This year we continued our research on QoE-based optimization routing for wireless mesh networks. The difficulties of the problem are analyzed and centralized and decentralized algorithms are proposed. The quality of the solution, the computational complexity of the proposed algorithm, and the fairness are our main concerns. Several centralized approximation algorithms have been already proposed in order to address the complexity and the quality of possible solutions. This year, we focused mainly on distributed algorithm to complement of the existing centralized algorithms. We propose decentralized heuristic algorithms based on the well-known Optimized Link-State Routing (OLSR) protocol. Control packets of OLSR are modified so as to be able to convey QoE-related information. The routing algorithm chooses the paths heuristically. After that, we studied message passing algorithms in order to find near optimal routing solutions in cooperative distributed networks. These algorithms have been published in [27], [13].

Sensors networks. In the literature, it is common to consider that sensor nodes in a clustered-based eventdriven Wireless Sensor Network (WSN) use a Carrier Sense Multiple Access (CSMA) protocol with a fixed transmission probability to control data transmission. However, due to the highly variable environment in these networks, a fixed transmission probability may lead to a significant amount of extra energy consumption. In view of this, three different transmission probability strategies for event-driven WSNs were studied in [51]: the optimal one, the "fixed" approach and a third "adaptive" method. As expected, the optimum strategy achieves the best results in terms of energy consumption but its implementation in a practical system is not feasible. The commonly used fixed transmission strategy (the probability for any node to attempt transmission is a constant) is the simplest approach but it does not adapt to changes in the system's conditions and achieves the worst performance. In the paper, we find that our proposed adaptive transmission strategy, where that probability is changed depending on specific conditions and in a very precise way, is pretty easy to implement and achieves results very close to the optimal method. The three strategies are analyzed in terms of energy consumption but also regarding the cluster formation latency. In [28], we also investigate cluster head selection schemes. Specifically, we consider two intelligent schemes based on the fuzzy C-means and k-medoids algorithms, and a random selection with no intelligence. We show that the use of intelligent schemes greatly improves the performance of the system, but their use entails higher complexity and some selection delay. The main performance metrics considered in this work are energy consumption, successful transmission probability and cluster formation latency. As an additional feature of this work, we study the effect of errors in the wireless channel and the impact on the performance of the system under the different considered transmission probability schemes.

Transmission delay, throughput and energy are also important criteria to consider in wireless sensor networks (WSNs). The IEEE 802.15.4 standard was conceived with the objective of reducing resource's consumption in both WSNs and Personal Area Networks (WPANs). In such networks, the slotted CSMA/CA still occupies a prominent place as a channel control access mechanism with its inherent simplicity and reduced complexity. In [26], we propose to introduce a network allocation vector (NAV) to reduce energy consumption and collisions in IEEE 802.15.4 networks. A Markov chain-based analytical model of the fragmentation mechanism, in a saturated traffic, is given as well as a model of the energy consumption using the NAV mechanism. The obtained results show that the fragmentation technique improves at the same time the throughput, the access delay and the bandwidth occupation. They also show that using the NAV allows reducing significantly the energy consumption when applying the fragmentation technique in slotted CSMA/CA under saturated traffic conditions.

#### 7.6. Optical Networks

#### Participants: Nicolás Jara, Gerardo Rubino

The rapid increase in demand for bandwidth in communication networks has caused a growth in the use of technologies based on WDM optical infrastructures. Nevertheless, in this last decade many researchers have recognized a "Capacity Crunch" associated with this technology, a transmission capacity limit on optical fibers, that is close to be reached pretty soon. This situation claims for an evolution on the currently used WDM optical architectures, in order to satisfy this relentless exponential growth in bandwidth demand. Following this trend, research started to examine in some detail specific aspects of the present functioning, and in particular, the way these networks are operated. Currently, optical networks are operated statically, but this is known to be inefficient in the usage of network resources, and with the previously mentioned upcoming risk of capacity collapse, it is of pressing matter to upgrade it. To this purpose, several proposals have been addressed and researched so far. Among these solutions, dynamic optical networks is the one closest to be implemented, but it has not been considered yet since the network cost savings are not enough to convince enterprises. This has been the focus of our research effort in the area.

The design of dynamic optical networks decomposes into different tasks, where the engineers must basically organize the way the main system's resources are used, minimizing the design and operation costs and respecting critical performance constraints. These tasks must guarantee certain level of quality of service (QoS) pre-established in the Service Level Agreement. In order to provide a proper quality of service measurement, we propose a new fast and accurate analytical method to evaluate the blocking probability that is at the heart of the path toward solving all the mentioned design problems. Blocking probability is the main QoS metric considered in the field. This work has been done in [20], where an analytical procedure has been proposed that combines efficiency and accuracy.

Next, the different tasks that must be addressed to find a good global design have been addressed in [19]. These are: which wavelength is going to be used by each user (the Wavelength Assignment Problem), how many wavelengths will be needed on each network link (the Wavelength Dimensioning Problem), and which set of paths enabling each network user to transmit (known as the Routing Problem) are to be established in order to minimize costs and to deal with link failures when the network is operating (this is the Fault Tolerance Problem). Two types of innovations and presented in this last paper. First, each of the problems receives a solution shown to be highly efficient. Second, and this is also new, we solve all the design problems simultaneously, using a single global algorithm (the usual way is to isolate them and to solve them one at a time, in a specific order). This work may provide a strategy to finally achieve sufficient cost savings, and thus, to contribute to make the decision to migrate from static to dynamic resource allocation easier. A preliminary version of a part of these results was presented previously in [44].

#### 7.7. Future networks and architectures

Participants: Jean-Michel Sanner, Hamza Ben Ammar, Louiza Yala, Yassine Hadjadj-Aoul, Gerardo Rubino

**SDN and NFV placement.** Mastering the increasing complexity of current and future networks, while reducing the operational and investments costs, is one of the major challenges faced by network operators (NOs). This explains in large part the recent enthusiasm of NOs towards Software Defined Networking (SDN) and Network Function Virtualization (NFV). Indeed, on the one hand, SDN makes it possible to get rid of the control plane distribution complexity, by centralizing it logically, while allowing its programmability. On the other hand, the NFV allows virtualizing the network functions, which considerably facilitates the deployment and the orchestration of the network resources. Providing a carrier grade network involves, however, several requirements such as providing a robust network meeting the constraints of the supported services. In order to achieve this objective, it is clearly necessary to scale network functions while placing them strategically in a way to guarantee the system's responsiveness.

The placement in TelCo networks are generally multi-objective and multi-constrained problems. The solutions proposed in the literature usually model the placement problem by providing a mixed integer linear program (MILP). Their performances are, however, quickly limited for large sized networks, due to the significant increase in the computational delays. In order to avoid the inherent complexity of optimal approaches and the

lack of flexibility of heuristics, we propose in [54] a genetic algorithm designed from the NSGA II framework that aims to deal with the controller placement problem. Genetic algorithms can be both multi-objective, multi-constraints and can be designed to be implemented in parallel. They constitute a real opportunity to find good solutions to this category of problems. Furthermore, the proposed algorithm can be easily adapted to manage dynamic placements scenarios. In [55], our main focus were devoted to maximize the clusters average connectivity and to balance the control's load between clusters, in a way to improve the networks' reliability.

We focus, in [60], on the problem of optimal computing resource allocation and placement for the provision of a virtualized Content Delivery Network (CDN) service over a telecom operator's Network Functions Virtualization (NFV) infrastructure. Starting from a Quality of Experience (QoE)-driven decision on the necessary amount of CPU resources to allocate in order to satisfy a virtual CDN deployment request with QoE guarantees, we address the problem of distributing these resources to virtual machines and placing the latter to physical hosts, optimizing for the conflicting objectives of management cost and service availability, while respecting physical capacity, availability and cost constraints. We present a multi-objective optimization problem formulation, and provide efficient algorithms to solve it by relaxing some of the original problem's assumptions. Numerical results demonstrate how our solutions address the trade-off between service availability and cost, and show the benefits of our approach compared with resource placement algorithms which do not take this trade-off into account.

**Real-time NFV placement in edge cloud.** Sometimes, the placement of NFV can not be planned in advance and therefore requires real-time placement as requests arrive. The placement is particularly challenging with the recent development of geographically distributed mini data centers, also referred to as cloudlets, at the edge of the network (i.e., typically at Points of Presence (PoPs) level). These edge data centers have rather small capacities in terms of storage, computing and networking resources, when compared with the huge centralized data centers deployed today.

All these radical changes in NOs' infrastructures raise many new issues (especially in terms of resource allocation), which so far have not been considered in the cloud literature. Traditionally, resources in cloud platforms are considered as to be infinite and request blocking is most of the time ignored when evaluating resources' allocation algorithms, precisely because of this infinite capacity assumption. However, if we assume that the NO's infrastructure will very likely be composed of small data centers with limited capacities, and deployed at the edge of network, the congestion of such a system may occur, notably if the demand is sufficiently high and exceeds what the infrastructure can handle at a given time.

We proposed in [57] an analytical model for the blocking analysis in a multidimensional cloud system, which was validated using discrete events' simulations. Besides, we conducted a comparative analysis of the most popular placement's strategies. The proposed model, as well as the comparative study, reveal practical insights into the performance evaluation of resource allocation and capacity planning for distributed edge cloud with limited capacities.

In [58] we set design principles of future distributed edge clouds in order to meet application requirements. We precisely introduce a costless distributed resource allocation algorithm, named *CLOSE*, which considers local information only. We compare via simulations the performance of *CLOSE* against those obtained by using mechanisms proposed in the literature, notably the Tricircle project within OpenStack. It turns out that the proposed distributed algorithm yields better performance while requiring less overhead.

In the context of the Open Network Automation Platform (ONAP), we develop in [56] a resource allocation strategy for deploying Virtualized Network Functions (VNFs) on distributed data centers. For this purpose, we rely on a three-level data center hierarchy exploiting co-location facilities available within Main and Core Central Offices. We precisely propose an active VNFs' placement strategy, which dynamically offloads requests on the basis of the load observed within a data center. We compare via simulations the performance of the proposed solution against mechanisms so far proposed in the literature, notably the centralized approach of the multi-site project within OpenStack, currently adopted by ONAP. Our algorithm yields better performance in terms of both data center occupancy and overhead. Furthermore, it allows extending the applicability of ONAP in the context of distributed cloud, without requiring any modification.

**Content Centric Networking.** Content-Centric Networking (CCN) has been proposed to address the challenges raised by the Internet usage evolution over the last years. One key feature provided by CCN to improve the efficiency of content delivery is the in-network caching, which has major impact on the system performance. In order to improve caching effectiveness in such systems, the study of the functioning of CCN innetwork storage must go deeper. In [39], we propose MACS, a Markov chain-based Approximation of CCN caching Systems. We start initially by modeling a single cache node. Then, we extend our model to the case of multiple nodes. A closed-form expression is then derived to define the cache hit probability of each content in the caching system. We compare the results of MACS to those obtained with simulations. The conducted experiments show clearly the accuracy of our model in estimating the cache hit performance of the system.

In [16], we present the design and implementation of a Content-Delivery-Network-as-a-Service (CDNaaS) architecture, which allows a telecom operator to open up its cloud infrastructure for content providers to deploy virtual CDN instances on demand, at regions where the operator has presence. Using northbound REST APIs, content providers can express performance requirements and demand specifications, which are translated into an appropriate service placement on the underlying cloud substrate. Our architecture is extensible, supporting various different CDN flavors, and, in turn, different schemes for cloud resource allocation and management. In order to decide on the latter in an optimal manner from an infrastructure cost and a service quality perspective, knowledge of the performance capabilities of the underlying technologies and computing resources is critical. Therefore, to gain insight which can be applied to the design of such mechanisms, but also with further implications on service pricing and SLA design, we carry out a measurement campaign to evaluate the capabilities of key enabling technologies for CDNaaS provision. In particular, we focus on virtualization and containerization technologies for implementing virtual CDN functions to deliver a generic HTTP service, as well as an HTTP video streaming one, empirically capturing the relationship between performance and service workload, both from a system operator and a user-centric viewpoints.

New tools for network design. In the efforts for designing future networks' topologies, the inclusion of dependability aspects has been recently enriched with finer criteria, and one relatively new family of metrics consider diameter-constrained parameters that capture more accurately reliability aspects of communication infrastructures. This is done by taking into account not only connectivity properties but also delays when nodes are connected. Paper [15] deals with factorization theory in diameter-constrained reliability, when terminal nodes are further required to be connected by d hops or fewer (d is a given strictly positive parameter of the metric, called its diameter). This metric was defined in 2001, inspired by delay-sensitive applications in telecommunications. Factorization theory is fundamental for classical network reliability evaluation, and today it is a mature area. However, its extension to the diameter-constrained context requires at least the recognition of irrelevant links, which is an open problem. In this paper, irrelevant links are efficiently determined in the most used case, where we consider the communication between a given pair of nodes in the network. The article also proposes a Factoring algorithm that includes the way series-parallels substructures can be handled.

**Quality of Experience activities.** We continue to develop tools for Quality of Experience assessment, and applications of this quantitative evaluation.

**Predicting time series.** For the future of the PSQA project, we intend to integrate the capability of *predicting* the Perceptual Quality and not only evaluating its current value. With this goal in mind, we explored this year the idea of combining a Reservoir Computing architecture (whose good performances have been reported many times, when used to predict sequences of numbers or of vectors) with Recurrent Random Neural Networks, that belong to a class of Neural Networks that have some nice properties. Both have been very successful in many applications. In [29] we propose a new model belonging to the first class, taking the structure of the second for its dynamics. The new model is called Echo State Queuing Network. The paper positions the model in the global Machine Learning area, and provides examples of its use and performances. We show on largely used benchmarks that it is a very accurate tool, and we illustrate how it compares with standard Reservoir Computing models. In [31] we presented some preliminary results to the Random Neural Network community.

**QoE and P2P design.** In [30] we describe a Peer-to-Peer (P2P) network that was designed to support Video on Demand (VoD) services. The network is based on a video-file sharing mechanism that classifies peers

according to the window (segment of the file) that they are downloading. This classification easily allows identifying peers that are able to share windows among them, so one of our major contributions is the definition of a mechanism that could be implemented to efficiently distribute video content in future 5G networks. Considering that cooperation among peers can be insufficient to guarantee an appropriate system performance, we also propose that this network must be assisted by upload bandwidth coming from servers; since these resources represent an extra cost to the service provider, especially in mobile networks, we complement our work by defining a scheme that efficiently allocates them only to those peers that are in windows with resources scarcity (we called it *prioritized windows distribution scheme*). On the basis of a fluid model and a Markov chain, we also develop a methodology that allows us to select the system parameters values (e.g., windows sizes or minimum servers upload bandwidth) that satisfy a set of Quality of Experience (QoE) parameters.

## 8. Bilateral Contracts and Grants with Industry

#### 8.1. Cifre contract on Device-Assisted Distributed Machine-Learning on Many Cores

#### Participants: Corentin Hardy, Bruno Sericola

This is a Cifre contract including a PhD thesis supervision (PhD of Corentin Hardy), done with Technicolor. The starting point of this thesis is to consider the possibility to deploy machine-learning algorithms over many cores, but out of the datacenter, on the devices (home-gateways) deployed by Technicolor in users' homes. In this device-assisted view, an initial processing step in the device may significantly reduce the burden on the datacenter back-end. Problems are numerous (power consumption, CPU power, network bandwidth and latency), but costs for the operator can be lowered and scale may bring some new level in data processing.

#### 8.2. Cifre contract on Throughput Prediction in Mobile Networks

#### Participants: Yann Busnel

This is a Cifre contract (2015-2018) including a PhD thesis supervision (PhD of Alassane Samba), done with Orange, on cooperation in statistical approaches for the prediction of throughput without history. Throughput has a strong impact on user experience in cellular networks. The ability to predict the throughput of a connection, before it starts, brings new possibilities, particularly to Internet service providers. They could adapt contents to the quality of service really reachable by users, in order to enhance their experience.

#### 8.3. Cifre contract on Mobile SDN architecture

Participants: Yassine Hadjadj-Aoul, César Viho

This is a Cifre contract (2015-2018) including a PhD thesis supervision (PhD of Imad Alawe), done with TDF, on the proposition of a scalable SDN-based mobile network architectures for the future 5G network.

## 8.4. Cifre contract on Personalization for Cognitive Autonomic Networks in 5G

#### Participants: César Viho

This is a Cifre contract (2017-2019) including a PhD thesis supervision (PhD of Illyyne Saffar), done with Nokia, on the proposition to use machine learning and data analytics to transform user and network data into actionable knowledge which in turn can be automatically exploited by Autonomic Networking approaches for cognitive self management of the 5G network.

#### 8.5. Bilateral Contract with Industry: ALSTOM-Inria Common Lab

Participants: Bruno Tuffin, Gerardo Rubino

Bruno Tuffin is the co-director of ALSTOM-Inria common Lab.

The group currently manages a project with ALSTOM on system availability simulation taking into account logistic constraints. Current ALSTOM Transport and Power contracts, especially service-level agreements, impose stringent system availability objectives. Non-adherence to the required performance levels often leads to penalties, and it is therefore critical to assess the corresponding risks already at a tender stage. The challenge is to achieve accurate results in a reasonable amount of time. Monte Carlo simulation provides estimates of the quantities it is desired to predict (e.g., availability). Since we deal with rare events, variance reduction techniques, specifically Importance Sampling (IS) here, is used. The goal of the project is to establish the feasibility of IS for solving problems relevant to ALSTOM and to develop the corresponding mathematical tools.

#### 8.6. Bilateral Contract with Industry: ADR Nokia Bell Labs

Participants: Yassine Hadjadj-Aoul, Gerardo Rubino

Gerardo Rubino is the coordinator of the reasearch action, named "Analytics and machine learning", with Nokia Bell Labs.

The objective is to carry out common research on an integrated framework for 5G, programmable networks, IoT and clouds that aims at statically and dynamically managing and optimizing the 5G infrastructure using, in particular, machine learning techniques.

## 9. Partnerships and Cooperations

#### 9.1. National Initiatives

#### 9.1.1. ANR

- Sofiène Jelassi is participating at 20% of his time to the IRT BCOM granted by the ANR.
- Yassine Hadjadj-Aoul is participating at 20% of his time to the IRT BCOM granted by the ANR.
- Yann Busnel is a member of the three following projects: SocioPlug granted by the ANR (ANR-13-INFR-0003), INSHARE granted by the ANR (ANR-15-CE19-0024) and BigClin granted by the LabEx CominLabs (ANR-10-LABX-07-01).

#### 9.1.2. IPL (Inria Project Lab) BetterNet

Yassine Hadjadj-Aoul, Gerardo Rubino and Bruno Tuffin are members of the IPL (Inria Project Lab) BetterNet: An Observatory to Measure and Improve Internet Service Access from User Experience, 2016-2020.

BetterNet aims at building and delivering a scientific and technical collaborative observatory to measure and improve the Internet service access as perceived by users. In this Inria Project Lab, we will propose new original user-centered measurement methods, which will associate social sciences to better understand Internet usage and the quality of services and networks. Our observatory can be defined as a vantage point, where: 1) tools, models and algorithms/heuristics will be provided to collect data, 2) acquired data will be analyzed, and shared appropriately with scientists, stakeholders and civil society, and 3) new value-added services will be proposed to end-users.

#### 9.2. European Initiatives

#### 9.2.1. Eurostars Camion Project

Participants: Yassine Hadjadj-Aoul

We were involved in a 30 months Eurostars European Project named Camion, which started on October 2014, aiming at offering cost-efficient, QoE-optimized content delivery, allowing for faster content access, as well as offline operation, while improving wireless network capacity and coverage. Camion is leaded by JCP-Connect, and the partners are a SME (FON) and our team. The project ended by June 2017.

#### 9.2.2. Collaborations in European Programs

#### 9.2.2.1. FINTEROP

Program: H2020-ICT-12-2015

Project acronym: F-Interop

Project title: FIRE+ online interoperability and performance test tools to support emerging technologies from research to standardization and market launch

Duration: November 2015 - October 2018

Coordinator: UPMC-LIP6

Other partners: 9 partners including (F. Sismondi and C. Viho (Dionysos); T. Watteyne (Eva))

Abstract: The goal of F-Interop is to extend FIRE+ with online interoperability and performance test tools supporting emerging IoT-related technologies from research to standardization and to market launch for the benefit of researchers, product development by SME, and standardization processes.

#### 9.2.3. Collaborations with Major European Organizations

Partner 1: Sapienza University of Rome, Italy.

We work with Nicoló Rivetti and Leonardo Querzoni on the analysis of stream processing systems.

#### 9.3. International Initiatives

#### 9.3.1. Inria International Partners

#### 9.3.1.1. Informal International Partners

- We maintain a strong line of collaborations with the Technical University Federico Santa María (UTFSM), Valparaíso, Chile. Over the years, this has taken different forms (associated team Manap, Stic AmSud project "AMMA", Stic AmSud project "DAT"). In 2017, we had a joint PhD work running (PhD of Nicolás Jara, to be defended at the beginning of next year), and a new joint PhD to be started in 2018 (PhD of Jonathan Olavarría). The first one is on optical network analysis and design, the second one on modeling evaluation techniques, with focus on Stochastic Activity Networks.
- We started a collaboration with the Faculty of Sciences of the university of the Republic, in Uruguay, on the application of mathematical modeling tools to a better understanding of a cognitive disease called semantic dementia. This involves Prof. Eduardo Mizraji and Jorge Graneri, PhD student, whose co-advisors are Prof. Mizraji and G. Rubino from Dionysos. Our contribution to this project is around the use of mathematical models, in particular around neural structures.

#### 9.3.2. Participation in Other International Programs

#### 9.3.2.1. International Initiatives

#### SM-HCD-HDD

Title: Statistical methods for highly complex and/or high dimensional data

International Partner (Institution - Laboratory - Researcher):

Universidad de la Republica Uruguay (Uruguay), Faculty of Sciences; Resp.: Ricardo Fraiman, Department of Mathematics

CNRS (France); Resp.: Catherine Aaron

Universidad Nacional del Litoral (Argentina); Resp.: Liliana Forzani

Duration: 3 years

#### Start year: 2016

In this project we work on specific statistical tools, mainly concerning predicting the behavior of time series. Our goal is to improve our tools for Perceptual Quality evaluation.

#### 9.3.2.2. International Initiatives

#### MOCQUASIN

Title: Monte Carlo and Quasi- Monte Carlo for rare event simulation

International Partner (Institution - Laboratory - Researcher):

Université de Montréal (Canada) - DIRO - Pierre L'Ecuyer

Duration: 3 years

Start year: 2013

See also: http://www.irisa.fr/dionysos/pages\_perso/tuffin/MOCQUASIN/

The goal of this team is to compute integrals, sums or to solve equations or optimization problems by means of Monte Carlo methods, which are statistical tools used when the models have a high complexity (for instance a large dimension). They are unavoidable tools in areas such as finance, electronics, seismology, computer science, engineering, physics, transport, biology, social sciences... Nonetheless, they have the reputation of being slow, i.e. to require a large computational time to reach a given precision. The goal of the project is to work on acceleration techniques, meaning methods allowing to reach the targeted precision in a shorter computational time. A typical framework is that of rare event simulation for which getting even only one occurrence of the event of interest could require a very long time. In this case, there are two main acceleration techniques: importance sampling and splitting, on which we work.

#### 9.4. International Research Visitors

#### 9.4.1. Visits of International Scientists

- Pierre L'Ecuyer holds an Inria International Chair, Nov. 2013- Oct. 2018.
- Marvin Nakayama (New Jersey Institute of Technology, NJ, USA) visited us 3 days in October to work on the estimation of quantiles in the case of rare events.
- Jonathan Olavarría, from UTFSM, Chile, from January to March (for two months), to work on stochastic models.
- Prof. Leslie Murray, from University of Rosario, Argentina (one month, February) to work on Monte Carlo techniques for rare event analysis.
- Jorge Graneri, from UDELAR, Uruguay (two months in the last quarter of the year, to work on biological applications).
- Prof. Claudio Risso, from UDELAR, Uruguay (two weeks in the last quarter of the year, to work on time series predictions).
- Prof. Gustavo Guerberoff, from UDELAR, Uruguay (two weeks in the last quarter of the year, to work on time series predictions).

## **10.** Dissemination

#### **10.1. Promoting Scientific Activities**

#### 10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

Pierre L'Ecuyer is member of the Steering Committee of MCQMC.

G. Rubino and B. Tuffin are members of the Steering Committee of the International Workshop on Rare Event Simulation (RESIM).

Y. Hadjadj-Aoul is co-chairing the Steering Committee of the International Conference on Information and Communication Technologies for Disaster Management (ICT-DM) from December 2016 and member of the steering committee since 2016.

Y. Hadjadj-Aoul is co-charing **ISNCC'2018**, "*The 5th International Symposium on Networks, Computers and Communications*" (Co-sponsored by IEEE), Roma, Italy (May 2018)

Yann Busnel has been member of the Organization Committee of AlgoTel 2017 (19èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications), which held in Quiberon in June 2017.

#### 10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

- Patrick Maillé and Bruno Tuffin were co-chairs of the International Workshop on Advanced Internet Charging and QoS technologies (ICQT'17), Tokyo, Japan, in November 2017
- Yassine Hadjadj-Aoul was co-chair of ISNCC'2017, "The 4th International Symposium on Networks, Computers and Communications" (Co-sponsored by IEEE), Marrakech, Morocco (May 2017)

#### 10.1.2.2. Member of the Conference Program Committees

Yann Busnel was a member of the Program Committee of the following events:

- NCA 2017: 16th IEEE International Symposium on Network Computing and Applications, Boston, USA, October 2017.
- CoRes 2017: 2èmes Rencontres Francophones sur la Conception de Protocoles, l'Évaluation de Performance et l'Expérimentation des Réseaux de Communication, Quiberon, France, May 2017.

Pierre L'Ecuyer was a member of the Program Committee of the following events:

- MCM'2017: Eleventh International Conference on Monte Carlo Methods and its Applications, Montreal, Canada, July 2017.
- ICORES 2017: International Conference on Operations Research and Enterprise Systems, Porto, Portugal, Feb. 2017.

Patrick Maillé was a member of the Program Committee of the following events:

- 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, Telecom ParisTech, Paris, France, 15th 19th May, 2017.
- NetEcon 2017: the 12th Workshop on the Economics of Networks, Systems and Computation, Cambridge, MA, USA, June 2017.

Bruno Sericola was a member of the Program Committee of the following event:

• ASMTA 2017: International Conference on Analytical and Stochastic Modelling Techniques and Applications, Newcastle-upon-Tyne, United Kingdom, 10-12 July 2017.

Gerardo Rubino was a member of the Program Committee of the following events:

- 11th International Conference on Monte Carlo Methods and Applications (MCM 2017), Montreal, Canada, July 3-7, 2017.
- XLIII Latin-American Conference in Computer Science (CLEI 2017), Córdoba, Argentina, September 4-8, 2017.
- IX International Congress of Computer Science and Telecommunications (COMTEL 2017), Lima, Peru, October 11-13, 2017.

Bruno Tuffin was a member of the Program Committee of the following events:

- 6th Workshop on Smart Data Pricing (SDP 2017), Workshop of IEEE INFOCOM 2017, Atlanta, GA, USA, May 2017.
- The Third International Symposium on Ubiquitous Networking (UNET 2017), Casablanca, Morocco, May 10-12, 2017.
- NetEcon 2017: the 12th Workshop on the Economics of Networks, Systems and Computation, Cambridge, MA, USA, June 2017.
- 11th International Conference on Monte Carlo Methods and Applications (MCM 2017), Montreal, Canada, July 3-7, 2017.
- 7th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH), Madrid, Spain, 29-31 July 2017.
- 14th International Conference on Economics of Grids, Clouds, Systems & Services (GECON'2017), Biarritz-Anglet-Bayonne, France, Sept. 19-21, 2017.
- The International Conference on Wireless Networks and Mobile Communications (WINCOM'17). Rabbat, Morocco, November 1-4, 2017.
- IEEE Globecom 2017, Singapore, Dec. 4-8, 2017
- 11th EAI International Conference on Performance Evaluation Methodologies and Tools (Value-Tools 2017), Venice, Italy, December 5-7, 2017

Yassine Hadjadj-Aoul was a member of the Program Committee of the following events:

- IEEE Globecom 2017, Singapore, Dec. 4-8, 2017
- IEEE ICC 2017, Paris, France, May. 21-25, 2017
- IEEE WCNC 2017, San Francisco, USA, March 19-22
- IEEE Symposium on Computers and Communications (ISCC 2017), Heraklion, Crete, Greece, July 3-6, 2017
- IEEE International Symposium on Networks, Computers and Communications (ISNCC 2017), Marrakech, Morroco, May 16-18, 2017

#### 10.1.2.3. Reviewer

Yann Busnel served as a reviewer for several major international conferences, such as DaWak 2017 (19th International Conference on Big Data Analytics and Knowledge Discovery).

Bruno Sericola served as a reviewer for several major international conferences.

Yassine Hadjadj-Aoul served as a reviewer for several major international conferences.

Gerardo Rubino served as a reviewer for several major international conferences, including those at which he served as a member of the Committee Program.

#### 10.1.3. Journal

#### 10.1.3.1. Member of the Editorial Boards

Bruno Tuffin is the Simulation Area Editor for Informs Journal on Computing.

Pierre L'Ecuyer is an associate editor for the following journals:

- ACM Transactions on Mathematical Software, since August 2004.
- Statistics and Computing (Springer-Verlag), since June 2003.
- International Transactions in Operational Research, since May 2007.

Bruno Tuffin is an associate editor for the following journal:

• ACM Transactions on Modeling and Computer Simulation, since November 2008.

Bruno Sericola is an associate editor for the following journals:

- International Journal of Stochastic Analysis, since April 2010.
- Performance Evaluation, since April 2015.

Bruno Sericola is Editor in Chief of the books series "Stochastic Models in Computer Science and Telecommunications Networks", ISTE/WILEY, since March 2015.

#### 10.1.3.2. Reviewer - Reviewing Activities

Yann Busnel served as a reviewer for several major international journals, such as TPDS (IEEE Transactions on Parallel and Distributed Systems).

In addition to the reports done during his associate editor and conference TPC member duties, Bruno Tuffin has reviewed papers in 2017 for IEEE JSAC, ACM TOIT, Telecom Policy, IEEE/ACM TON.

Bruno Sericola served as a reviewer for several major international journals.

Yassine Hadjadj-Aoul served as a reviewer for several major international journals, such as TVT (IEEE Transaction on Vehicular Technology) and IEEE JSAC.

César Viho reviewed papers for the journals IEEE Transaction on Wireless Communication, IEEE Transactions on Vehicular Communications, IEEE Communications Magazine, and for the following international conferences: IWCNC, Globecom, and CCNC.

Gerardo Rubino served as a reviewer for several major international journals.

#### 10.1.4. Invited Talks

Yann Busnel made several invited and keynote talks in 2017:

- Analyse et traitement de flux de données à large échelle, Invited talk at Journées non thématiques RESCOM 2017, Nice, France, January 2017.
- *Analyse et traitement de flux de données à large échelle* Invited talk at Journées ARC du GRD MACS 2017, Paris, France, May 2017.
- Ordonnancement dynamique pour un équilibrage de charge quasi-optimal dans les systèmes de traitement de flux, Plenary talk at AlgoTel 2017, Quiberon, France, June 2017.
- Comment créer un cloud social sécurisé pour ses données ? Le projet SocioPlug ! invited talk at 23ème Technoférence du Pôle Images & Réseaux, Nantes, France, December 2017.

B. Tuffin gave a keynote talk "Network Neutrality: Modeling, Challenges, and its Impact on Clouds" at the 14th International Conference on Economics of Grids, Clouds, Systems & Services (GECON 2017), Keynote talk, Bayonne Anglet, France, September 19-21, 2017.

B. Tuffin gave the following seminar presentation:

- B. Tuffin. Neutralité du Net: introduction, modélisation et défis. ENS Rennes, Feb. 7, 2017.
- B. Tuffin. La fausse neutralité du net ? ENS-ENSSIB, Lyon, Mars 2017.

Yassine Hadjadj-Aoul led a panel on "Intelligence defined network for future smart cities" during the 4th International Symposium on Networks, Computers and Communications (ISNCC 2017).

Raymond Marie gave a seminar at the Polytechnic University of Hong Kong during a two-week stay as a guest.

G. Rubino made several invited and keynote talks in 2017. Two around Monte Carlo techniques:

- "Rare events in simulation: issues and techniques", plenary talk, for the micro-simulation community (see [36]).
- "Dependability Analysis through Monte Carlo Methods: The Case of Rare Events", a tutorial (see [33]).

and two around transient analysis of Markovian processes:

- "On the derivation of closed-form expressions of the solutions to (possibly infinite) some simple linear systems of ODEs", keynote for researchers in dynamical systems and differential equations (see [35]).
- "New results on the transient analysis of some fundamental queuing systems", keynote oriented to modeling (see [34]).

#### 10.1.5. Leadership within the Scientific Community

Yann Busnel is a member of the CSV (the technical committee) of the Images and Networks Cluster of Brittany, France.

Yann Busnel is a member of the Steering Committee of the RESCOM research group at GDR CNRS RSD.

Yassine Hadjadj-Aoul is a founding member of Special Interests Group "IEEE Sig on Big Data with Computational Intelligence", under the IEEE COMSOC Big Data TC (since June 2017).

Yassine Hadjadj-Aoul is a member of the GT ARC (Automatique et Réseaux de Communication) scientific committee (since Nov. 2017)

Gerardo Rubino is one of the three French representatives at the Scientific Committee of the IFCAM (Indo-French Centre for Applied Mathematics), managing the cooperation in mathematics of the two countries, and federating at the French side, among several other participants, Inria and CNRS.

Gerardo Rubino is a member of the CSV (the technical committee) of the Images and Networks Cluster of Brittany, France.

#### 10.1.6. Scientific Expertise

César Viho has reviewed project proposals for the ANR and for CIFRE contracts for the ANRT.

Gerardo Rubino has participated as an expert for several institutions abroad, and for different tasks (recruitments, promotions, prizes): the University of Vienna, Austria, the University of New South Wales, Australia, the UTFSM, Chile.

#### 10.1.7. Research Administration

- Bruno Tuffin is the co-director of the common lab ALSTOM-Inria since 2014.
- Bruno Tuffin was a member of Inria-MITACS selection committee.
- Bruno Tuffin was a member of Inria Rennes' "Inria delegation" selection committee.
- Bruno Tuffin was a member of Inria Rennes' post-doc selection committee.
- Yann Busnel is head of "Network System, Cybersecurity and Digital law" Research Department at IMT Atlantique.
- Yann Busnel is member of Development Council of Computer Sciences Master at University of Nantes.
- Bruno Sericola is responsible for the Inria Rennes-Bretagne Atlantique budget.
- Bruno Sericola is the leader of the research group MAPI (Math Appli Pour l'Info) the goal of which is to improve the collaboration between computer scientists and mathematicians.
- César Viho is director of the MathSTIC (Mathematics, Electronics and Computer Sciences) doctoral school in charge of managing the recruitment of around 1100 PhD students and their activities during their doctorate, in all the concerned areas of the UBL (Université Bretagne Loire).

#### 10.2. Teaching - Supervision - Juries

#### 10.2.1. Teaching

Master: Bruno Tuffin, MEPS (performance evaluation), 35h, M1, Univ Rennes, France

Master: Bruno Tuffin, GTA (Game Theory and Applications), 15h, M2, Univ Rennes, France Master: Patrick Maillé, GTA (Game Theory and Applications), 15h, M2, Univ Rennes, France Master: Patrick Maillé, Simulation and queuing theory, 25h, M2, IMT Atlantique, France Licence: Patrick Maillé, Techniques and models in networks, 20h, L3, IMT Atlantique, France Master: Patrick Maillé, Performance Evaluation, 30h, M1, IMT Atlantique, France Licence: Yann Busnel, Introduction to Network, 15h, 1st year ENS Rennes, France Master: Yann Busnel, Big Data and Stream Processing, 9h, IMT Atlantique, Rennes, France Master: Bruno Sericola, Mathematics, 12h, M2, Istic/University of Rennes 1, France. Master: Bruno Sericola, Logistic and performance, 12h, M2, Faculté de sciences économiques, Univ of Rennes 1, France Master: Bruno Sericola, MEPS (performance evaluation), 36h, M1, Univ Rennes, France Master pro 2nd year: Yassine Hadjadj-Aoul, Multimedia streaming over IP (MMR), 48 hours, Esir/University of Rennes 1, France Master pro 2nd year: Yassine Hadjadj-Aoul, Multimedia services in IP networks (RSM), 29 hours, Esir/University of Rennes 1, France Master pro 2nd year: Yassine Hadjadj-Aoul, Software Defined Networks, 6 hours, Istic/University of Rennes 1, France Master 2nd year: Yassine Hadjadj-Aoul, Video streaming over IP, 8 hours, Istic/University of Rennes 1, France Master: Yassine Hadjadj-Aoul, Introduction to networking (IR), 26 hours, Esir/University of Rennes 1, France Master: Yassine Hadjadj-Aoul, Mobile and wireless networking (RMOB), 20 hours, Esir/University of Rennes 1, France Master 2nd year: Yassine Hadjadj-Aoul, Overview of IoT technologies: focus on LPWAN, 2 hours, **INSA**, France Master pro 2nd year: Sofiéne Jelassi, Supervision of heterogeneous networks, 32 hours, Istic/University of Rennes 1, France Master pro 2nd year: Sofiéne Jelassi, Cloud & SDN virtualization, 32 hours, Istic/University of Rennes 1, France Master pro 2nd year: Sofiéne Jelassi, Multimedia networks, 32 hours, Istic/University of Rennes 1, France Master 2nd year: Sofiéne Jelassi, Software defined networking, 6 hours, Istic/University of Rennes 1, France Master M1: César Viho, Networks: from Services to protocols, 36 hours, Istic/University of Rennes 1, France Master M2: César Viho, Algorithms on graphs, 40 hours, Istic/University of Rennes 1, France Bachelor L2: César Viho, Network architecture and components, 16 hours, Istic/University of Rennes 1, France Supelec Rennes 3rd year: Gerardo Rubino, Dependability Analysis, 15 hours. UDELAR, Uruguay: Gerardo Rubino, post-graduate course on dependability, 21 hours. 10.2.2. Supervision PhD in progress: Ajit Rai, "Availability prediction with logistics", started in May 2015; advisors: B. Tuffin & G. Rubino, University Rennes 1.

Joshua Peignier, Estelle Varloot. "Game-theoretic tools to analyze classical vs collaborative economies", Project Master R&I, 2016-2017. Advisors: P. Maillé and B. Tuffin.

PhD in progress: Alassane Samba, "Technologies Big Data et modèles prédictifs pour la QoS des réseaux", IMT Atlantique. Advisors: Y. Busnel, G. Simon and P. Dooze (Inria). Defense in 2018.

PhD in progress: Richard Westerlynck, "Analyse répartie et extraction de tendances à grande échelle pour les données massives en santé", IMT Atlantique. Advisors: Y. Busnel and M. Cuggia (PUPH, CHU Rennes). Defense in 2020.

PhD in progress: Vasile Cazacu, "Calcul distribué pour la fouille de données cliniques", IMT Atlantique. Advisors: E. Anceaume (CNRS Rennes), Y. Busnel and M. Cuggia (PUPH, CHU Rennes). Defense in 2020.

PhD in progress: Corentin Hardy, "Device-Assisted Distributed Machine-Learning on Many Cores", started in April 2016; advisors: Bruno Sericola and Erwan Le Merrer from Technicolor, University Rennes 1.

PhD in progress: Yves Mocquard, "Analyse de flux de données massifs dans les systèmes distribués large échelle", started on September 2015; advisors: Bruno Sericola and Emmanuelle Anceaume from team Cidre, University Rennes 1.

PhD in progress: Ali Hodroj, "Enhancing content delivery to multi-homed users in broadband mobile networks", started in November 2015; advisors: Bruno Sericola, Marc Ibrahim and Yassine Hadjadj-Aoul, University Rennes 1 and St Joseph University of Beyrouth.

PhD in progress: Jean-Michel Sanner; Cifre Grant, Orange Labs, "SDN technologies for network services performances adaptation of carriers networks"; started on January 2013;

PhD in progress: Hamza Ben Ammar, "Socially-aware network and cache resources optimization for efficient media content delivery in Content Centric Networks", started in October 2015; advisors: Yassine Hadjadj-Aoul, Adlen Ksentini and Soraya Ait Chellouche, University Rennes 1.

PhD in progress: Imad Alawe, "Mobile SDN architecture", started in October 2015; advisors: César Viho, Yassine Hadjadj-Aoul, University Rennes 1, Philippe Bertin, B-COM and Davy Darche, TDF.

PhD in progress: Jean-Michel Sanner; Cifre Grant, Orange Labs, "SDN technologies for network services performances adaptation of carriers networks"; started on January 2013; Advisors: Y. Hadjadj-Aoul and G. Rubino; University Rennes 1.

PhD in progress: Imane Taibi, "Big data analysis for network monitoring and troubleshooting"; started on Nov. 2017; Advisors: G. Rubino, Inria, and Yassine Hadjadj-Aoul, University Rennes 1, and Chadi Ibrahim, Inria.

PhD in progress: Mohamed Rahali, "Machine learning-based monitoring and management for hybride SDN networks"; started on Oct. 2017; Advisors: G. Rubino, Inria, and Sofiène Jelassi, University of Rennes 1.

PhD in progress: Nicolás Jara, "Fault tolerant design of dynamic WDM optical networks", Technical University Federico Santa María (UTFSM) and university of Rennes 1, France. Advisors: R. Vallejos (Chile) and G. Rubino (France). Defense in 2018.

PhD in progress: Laura Aspirot, "Fluid Approximations for Stochastic Telecommunication Models", University of the Republic, Uruguay. Advisors: E. Mordecki (Uruguay) and G. Rubino (France). Defense in 2018.

PhD in progress: Jorge Graneri, "Mathematical Models for Semantic Memory", University of the Republic, Uruguay. Advisors: E. Mizraji (Uruguay) and G. Rubino (France). Started end 2016.

#### 10.2.3. Juries

Bruno Tuffin was a member of the following PhD defense committee:

• Kodjo Séna Apeke. Modélisation ubiquiste pour l'interaction d'échelles. Application à la prédiction de la réponse d'une tumeur sous traitement en radiothérapie. Université de Bretagne Occidentale, 2017.

Bruno Sericola was member of the final selecting board for the recruitment of CNRS researchers in 2017.

Yassine Hadjadj-Aoul was a member of the PhD defense committee of Souheir Eido, IMT Atlantique, Brest (2017)

Yassine Hadjadj-Aoul and Gerardo Rubino were members of the PhD jury of Yue Li. Title: "Elaboration d'une architecture réseau unifiée, ouverte et flexible". Defense: September 29, 2017.

César Viho was a member of the following juries:

- Recruitment of young graduate scientists and senior researchers at Inria.
- Recruitment of young Associate Professors and senior Professors at ISTIC-Université Rennes 1.

#### **10.3.** Popularization

- B. Tuffin. La simulation de Monte-Carlo. Interstices. 2017. https://interstices.info/jcms/int\_69164/ la-simulation-de-monte-carlo
- G. Rubino makes regular presentations to high school students about the research work in general, and specific technical topics in particular. Current talks:
  - Randomness as a tool
  - Internet as a research problem
  - Great challenges in maths: the Riemann Hypothesis
  - Great challenges in math/computer science: the "P versus NP" problem

## 11. Bibliography

#### Major publications by the team in recent years

- [1] G. RUBINO, B. TUFFIN (editors). *Rare Event Simulation using Monte Carlo Methods*, John Wiley & Sons, 2009.
- [2] N. BOUABDALLAH, A.-L. BEYLOT, E. DOTARO, G. PUJOLLE. Resolving the Fairness Issues in Bus-Based Optical Access Networks, in "IEEE Journal on Selected Areas in Communications", 2005, vol. 23, n<sup>o</sup> 8, p. 1444–1457.
- [3] Y. HADJADJ-AOUL, T. TALEB.An adaptive fuzzy-based CAC scheme for uplink and downlink congestion control in converged IP and DVB-S2 networks, in "IEEE Transactions on Wireless Communications", Feb. 2009, vol. 8, n<sup>o</sup> 2, p. 816–825.
- [4] Y. HAYEL, D. ROS, B. TUFFIN. Less-than-Best-Effort Services: Pricing and Scheduling, in "23rd IEEE Infocom Conference", Hong-Kong, China, March 2004.
- [5] P. L'ECUYER. Randomized Quasi-Monte Carlo: An Introduction for Practitioners, in "12th International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing (MCQMC 2016)", Stanford, United States, August 2016, https://hal.inria.fr/hal-01561550.
- [6] P. LEGUESDRON, J. PELLAUMAIL, G. RUBINO, B. SERICOLA. Transient analysis of the M/M/1 queue, in "Advances in Applied Probability", September 1993, vol. 25, n<sup>o</sup> 3, p. 702–713.
- [7] P. MAILLÉ, B. TUFFIN. *Telecommunication Network Economics: From Theory to Applications*, Cambridge University Press, 2014, 288, https://hal.inria.fr/hal-00908598.

- [8] H. NABLI, B. SERICOLA. Performability analysis: a new algorithm, in "IEEE Transactions on Computers", 1996, vol. 45, n<sup>o</sup> 4, p. 491–494.
- [9] A. NAFAA, A. KSENTINI. On Sustained QoS Guarantees in Operated IEEE 802.11 Wireless LANs, in "IEEE Transactions on Parallel and Distributed Systems", 2008, vol. 19, n<sup>o</sup> 8, p. 1020–1033.
- [10] G. RUBINO, B. SERICOLA. *A finite characterization of weak lumpable Markov processes. Part II: The continuous time case*, in "Stochastic Processes and their Applications", 1993, vol. 45, p. 115–126.
- [11] B. TUFFIN. Bounded Normal Approximation in Highly Reliable Markovian Systems, in "Journal of Applied Probability", 1999, vol. 36, n<sup>o</sup> 4.

#### **Publications of the year**

#### **Doctoral Dissertations and Habilitation Theses**

- [12] Y. LI. Edge computing-based access network selection for heterogeneous wireless networks, Université Rennes 1, September 2017, https://tel.archives-ouvertes.fr/tel-01661436.
- [13] Q. PHAM TRAN ANH.Algorithms and optimization for quality of experience aware routing in wireless networks : from centralized solutions, Université Rennes 1, January 2017, https://tel.archives-ouvertes.fr/tel-01488283.

#### **Articles in International Peer-Reviewed Journal**

- [14] E. ANCEAUME, Y. BUSNEL.Lightweight Metric Computation for Distributed Massive Data Streams, in "Transactions on Large-Scale Data- and Knowledge-Centered Systems", April 2017, vol. 10430, n<sup>o</sup> 33, p. 1–39 [DOI: 10.1007/978-3-662-55696-2\_1], https://hal.archives-ouvertes.fr/hal-01634353.
- [15] E. CANALE, P. ROMERO, G. RUBINO. Factorization and Exact Evaluation of the Source-Terminal Diameter-Constrained Reliability, in "Networks", 2017, vol. 70, n<sup>o</sup> 4, p. 283-291 [DOI: 10.1002/NET.21780], https:// hal.inria.fr/hal-01663508.
- [16] P. FRANGOUDIS, L. YALA, A. KSENTINI.CDN-as-a-Service Provision over a Telecom Operator's Cloud, in "IEEE Transactions on Network and Service Management (TNSM)", June 2017, vol. 14, n<sup>o</sup> 3, p. 702-716 [DOI: 10.1109/TNSM.2017.2710300], https://hal.inria.fr/hal-01659931.
- [17] S. GAMBOA, A. PELOV, P. MAILLÉ, X. LAGRANGE, N. MONTAVONT. Reducing the energy footprint of cellular networks with delay-tolerant users, in "IEEE Systems Journal", June 2017, vol. 11, n<sup>o</sup> 2, p. 729 - 739 [DOI: 10.1109/JSYST.2015.2476376], https://hal.archives-ouvertes.fr/hal-01611343.
- [18] E. GOURDIN, P. MAILLÉ, G. SIMON, B. TUFFIN. The Economics of CDNs and Their Impact on on Service Fairness, in "IEEE Transactions on Network and Service Management", 2017, vol. 14, n<sup>o</sup> 1, p. 22-33, https:// hal.inria.fr/hal-01398923.
- [19] N. JARA, R. VALLEJOS, G. RUBINO.A Method for Joint Routing, Wavelength Dimensioning and Fault Tolerance for any set of simultaneous failures on Dynamic WDM Optical Networks, in "Optical Fiber Technology", 2017, vol. 38, n<sup>o</sup> Supplement C, p. 30-40 [DOI: 10.1016/J.YOFTE.2017.08.001], https:// hal.inria.fr/hal-01663501.

- [20] N. JARA, R. VALLEJOS, G. RUBINO. Blocking Evaluation and Wavelength Dimensioning of Dynamic WDM Networks without Wavelength Conversion, in "Journal of optical communications and networking", August 2017, vol. 9, n<sup>o</sup> 8, p. 625-634 [DOI : 10.1364/JOCN.9.000625], https://hal.inria.fr/hal-01663506.
- [21] P. L'ECUYER, P. MAILLÉ, N. STIER-MOSES, B. TUFFIN. Non-Neutrality of Search Engines and its Impact on Innovation, in "Internet Technology Letters", 2017, https://hal.inria.fr/hal-01634370.
- [22] P. L'ECUYER, D. MUNGER, B. ORESHKIN, R. SIMARD.Random Numbers for Parallel Computers: Requirements and Methods, in "Mathematics and Computers in Simulation", 2017, vol. 135, p. 3-17, https://hal.inria. fr/hal-01399537.
- [23] P. MAILLÉ, B. TUFFIN. Auctions for online ad space among advertisers sensitive to both views and clicks, in "Electronic Commerce Research", 2017, https://hal.inria.fr/hal-01241308.
- [24] P. MAILLÉ, B. TUFFIN. Preventing competition using side payments: when non-neutrality creates barriers to entry, in "Netnomics", 2017, vol. 18, n<sup>o</sup> 1, p. 3-22 [DOI: 10.1007/s11066-016-9110-6], https://hal.inria. fr/hal-01398913.
- [25] F. MESSAOUDI, A. KSENTINI, G. SIMON, P. BERTIN. On using Edge Computing for computation offloading in mobile network, in "ACM Transactions on Multimedia Computing, Communications and Applications", October 2017, vol. 13, n<sup>0</sup> 4, p. 1 - 28 [DOI : 10.1145/3115934], https://hal.inria.fr/hal-01661889.
- [26] A. MOULOUD, D. AISSANI, Y. HADJADJ-AOUL. Towards Bandwidth and Energy Optimization in IEEE 802.15.4 Wireless Sensor Networks, in "Springer Computing", October 2017, p. 1-24 [DOI: 10.1007/s00607-017-0583-4], https://hal.inria.fr/hal-01657663.
- [27] Q. PHAM TRAN ANH, K. PIAMRAT, K. D. SINGH, C. VIHO. Video Streaming over Ad-hoc Networks: a QoE-based Optimal Routing Solution, in "IEEE Transactions on Vehicular Technology", February 2017, vol. 66, nº 2 [DOI : 10.1109/TVT.2016.2552041], https://hal.archives-ouvertes.fr/hal-01340989.
- [28] E. ROMO MONTIEL, M. RIVERO-ANGELES, G. RUBINO, H. MOLINA-LOZANO, R. MENCHACA-MÉNDEZ, R. MENCHACA-MÉNDEZ. Performance Analysis of Cluster Formation in Wireless Sensor Networks, in "Sensors", December 2017, vol. 17, n<sup>o</sup> 12, p. 1-49 [DOI: 10.3390/s17122902], https://hal.inria. fr/hal-01665857.
- [29] G. RUBINO, S. BASTERRECH. Echo State Queueing Networks: a combination of Reservoir Computing and Random Neural Networks, in "Probability in the Engineering and Informational Sciences", 2017, vol. 31, n<sup>o</sup> 4, p. 457-476 [DOI: 10.1017/S0269964817000110], https://hal.inria.fr/hal-01663499.
- [30] N. TORRES-CRUZ, M. RIVERO-ANGELES, G. RUBINO, R. MENCHACA-MÉNDEZ, R. MENCHACA-MÉNDEZ.A window-based, server-assisted P2P network for VoD services with QoE guarantees, in "Mobile Information Systems", March 2017, vol. 2017, n<sup>o</sup> Article ID 2084684, p. 1-16 [DOI: 10.1155/2017/2084684], https://hal.inria.fr/hal-01665853.

#### **Invited Conferences**

[31] S. BASTERRECH, G. RUBINO.A G-Network as a Reservoir Computing Model, in "First G-Networks Workshop", Nice, France, January 2017, https://hal.inria.fr/hal-01658101.

- [32] P. L'ECUYER. *History of uniform random number generation*, in "WSC 2017 Winter Simulation Conference", Las Vegas, United States, December 2017, https://hal.inria.fr/hal-01561551.
- [33] G. RUBINO. Dependability Analysis through Monte Carlo Methods: The Case of Rare Events Tutorial, in "ICDT 2017 - The 12th International Conference on Digital Telecommunications", Venezia, Italy, April 2017, https://hal.inria.fr/hal-01658088.
- [34] G. RUBINO.New results on the transient analysis of some fundamental queuing systems Keynote presentation, in "ACMPT 2017 - Analytical and Computational Methods in Probability Theory and its Applications", Moscow, Russia, A. LEBEDEV (editor), Analytical and Computational Methods in Probability Theory, Lomonosov Moscow State University (MSU), October 2017, https://hal.inria.fr/hal-01657448.
- [35] G. RUBINO.On the derivation of closed-form expressions of the solutions to (possibly infinite) some simple linear systems of ODEs - Keynote presentation, in "CDEDS 2017 - Conference on Differential Equations and Dynamical Systems", Suzhou, China, September 2017, https://hal.inria.fr/hal-01658064.
- [36] G. RUBINO.*Rare events in simulation: issues and techniques Plenary speaker*, in "IMA 2017 the 6th World Congress of the Microsimulation Association", Torino, Italy, June 2017, https://hal.inria.fr/hal-01658082.

#### **International Conferences with Proceedings**

- [37] I. ALAWE, Y. HADJADJ-AOUL, A. KSENTINI, P. BERTIN, D. DARCHE. On the scalability of 5G Core network: the AMF case, in "CCNC 2018 - IEEE Consumer Communications and Networking Conference", Las Vegas, United States, January 2018, p. 1-6, https://hal.inria.fr/hal-01657667.
- [38] I. ALAWE, A. KSENTINI, Y. HADJADJ-AOUL, P. BERTIN, A. KERBELLEC. On Evaluating Different Trends for Virtualized and SDN-ready Mobile Network, in "CloudNet 2017 - 6th IEEE International Conference on Cloud Networking", Prague, Czech Republic, Cloud Networking (CloudNet), 2017 IEEE 6th International Conference on, IEEE, September 2017, p. 1-6 [DOI: 10.1109/CLOUDNET.2017.8071534], https://hal.inria. fr/hal-01657671.
- [39] H. B. AMMAR, S. A. CHELLOUCHE, Y. HADJADJ-AOUL. *Markov Chain-based Approximation of CCN Caching Systems*, in "ISCC 2017 IEEE Symposium on Computers and Communications", Heraklion, Greece, IEEE, July 2017, p. 1-6 [DOI: 10.1109/ISCC.2017.8024551], https://hal.inria.fr/hal-01657679.
- [40] Z. BOTEV, P. L'ECUYER. Accurate computation of the right tail of the sum of dependent log-normal variates, in "WSC 2017 - Winter Simulation Conference", Las Vegas, United States, December 2017, https://hal.inria. fr/hal-01561552.
- [41] G. BOUZILLÉ, R. WESTERLYNCK, G. DEFOSSEZ, D. BOUSLIMI, S. BAYAT, C. RIOU, Y. BUSNEL, C. LE GUILLOU, J.-M. CAUVIN, C. JACQUELINET, P. PLADYS, E. OGER, E. STINDEL, P. INGRAND, G. COATRIEUX, M. CUGGIA. Sharing health big data for research - A design by use cases: the INSHARE platform approach, in "The 16th World Congress on Medical and Health Informatics (MedInfo2017)", Hangzhou, China, The 16th World Congress on Medical and Health Informatics (MedInfo2017), August 2017, https://hal.archives-ouvertes.fr/hal-01535703.
- [42] P. W. GLYNN, M. K. NAKAYAMA, B. TUFFIN. On the estimation of the mean time to failure by simulation, in "2017 Winter Simulation Conference", Las Vegas, United States, December 2017, https://hal.inria.fr/hal-01551078.
#### [43] Best Paper

C. HARDY, E. LE MERRER, B. SERICOLA. *Distributed deep learning on edge-devices: feasibility via adaptive compression*, in "IEEE NCA 2017 - 16th IEEE International Symposium on Network Computing and Applications", Boston, United States, October 2017, This article has received the Best Paper Award, https://hal.inria.fr/hal-01650936.

- [44] N. JARA, R. VALLEJOS, G. RUBINO. Alternate Paths for Multiple Fault Tolerance on Dynamic WDM Optical Networks, in "HPSR 2017 - 18th IEEE International Conference on High Performance Switching and Routing", Campinas, Brazil, IEEE, June 2017, p. 1-6 [DOI: 10.1109/HPSR.2017.7968674], https:// hal.inria.fr/hal-01663507.
- [45] Y. LI, P. FRANGOUDIS, Y. HADJADJ-AOUL, P. BERTIN. A Mobile Edge Computing-Assisted Video Delivery Architecture for Wireless Heterogeneous Networks, in "ISCC 2017 - IEEE Symposium on Computers and Communications", Heraklion, Greece, IEEE, July 2017, p. 1-6 [DOI : 10.1109/ISCC.2017.8024583], https://hal.inria.fr/hal-01657678.
- [46] P. MAILLÉ, B. TUFFIN. Enforcing free roaming among UE countries: an economic analysis, in "13th International Conference on Network and Service Management (CNSM)", Tokyo, Japan, November 2017, https://hal.inria.fr/hal-01428920.
- [47] P. MAILLÉ, B. TUFFIN. How does imposing free roaming in EU impact users and ISPs' relations?, in "8th International Conference Network of the Future", London, United Kingdom, November 2017, https://hal.inria. fr/hal-01493911.
- [48] P. MAILLÉ, B. TUFFIN. Non-Neutrality Pushed by Big Content Providers, in "14th International Conference on Economics of Grids, Clouds, Systems, and Services (GE- CON'2017)", Biarritz, France, September 2017, https://hal.inria.fr/hal-01398926.
- [49] F. MESSAOUDI, A. KSENTINI, P. BERTIN. On using Edge Computing for computation offloading in mobile network, in "Globecom 2017 - IEEE Global Communications Conference", Singapore, Singapore, December 2017, https://hal.inria.fr/hal-01661885.
- [50] Y. MOCQUARD, B. SERICOLA, E. ANCEAUME. Probabilistic Analysis of Counting Protocols in Largescale Asynchronous and Anonymous Systems, in "Proceedings of the 16th IEEE International Conference on Network Computing and Applications", Boston, United States, October 2017, https://hal.archives-ouvertes.fr/ hal-01580275.
- [51] M. RIVERO-ANGELES, G. RUBINO. Transmission Probability Strategies for Cluster-based Event-Driven Wireless Sensor Networks, in "CyberC 2017 - International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery", Nanjing, China, October 2017, p. 1-4, https://hal.inria.fr/hal-01665855.
- [52] N. RIVETTI, Y. BUSNEL, L. QUERZONI. Délestage avisé dans les systèmes de traitement de flux, in "AL-GOTEL 2017 19èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Quiberon, France, May 2017, https://hal.archives-ouvertes.fr/hal-01519427.

- [53] A. SAMBA, Y. BUSNEL, A. BLANC, P. DOOZE, G. SIMON.*Instantaneous Throughput Prediction in Cellular Networks: Which Information Is Needed?*, in "2017 IFIP/IEEE International Symposium on Integrated Network Management (IM)", Lisbonne, Portugal, 2017 IFIP/IEEE International Symposium on Integrated Network Management (IM), IEEE, May 2017, https://hal.archives-ouvertes.fr/hal-01535696.
- [54] J.-M. SANNER, Y. HADJADJ-AOUL, M. OUZZIF, G. RUBINO. An evolutionary controllers' placement algorithm for reliable SDN networks, in "ManSDN/NF 2017 - 4th International Workshop on Management of SDN and NFV Systems", Tokyo, Japan, November 2017, p. 1-6, https://hal.inria.fr/hal-01657698.
- [55] J.-M. SANNER, M. OUZZIF, Y. HADJADJ-AOUL.Multi-objective evolutionary algorithms for optimized placements of SDN controllers, VNFs and VNFs chains in new Telecom infrastructures, in "SDN Day 2017 -Software Defined Networks", Paris, France, November 2017, https://hal.inria.fr/hal-01657689.
- [56] F. SLIM, F. GUILLEMIN, A. GRAVEY, Y. HADJADJ-AOUL. *Towards a Dynamic Adaptive Placement of Virtual Network Functions under ONAP*, in "Third International NFV-SDN'17-O4SDI Workshop on Orchestration for Software-Defined Infrastructures)", Berlin, Germany, November 2017, p. 1-6, https://hal.inria.fr/hal-01657705.
- [57] F. SLIM, F. GUILLEMIN, Y. HADJADJ-AOUL. On Virtual Network Functions' Placement in Future Distributed Edge Cloud, in "CloudNet 2017 - IEEE 6th International Conference on Cloud Networking", Prague, Czech Republic, Cloud Networking (CloudNet), 2017 IEEE 6th International Conference on, IEEE, September 2017, p. 1-4 [DOI: 10.1109/CLOUDNET.2017.8071524], https://hal.inria.fr/hal-01657676.
- [58] F. SLIM, F. GUILLEMIN, Y. HADJADJ-AOUL.CLOSE: A Costless Service Offloading Strategy for Distributed Edge Cloud, in "GTCC 2017 - IEEE Workshop on Game Theory in Computer Communications", Las Vegas, United States, January 2018, https://hal.inria.fr/hal-01661624.
- [59] B. TUFFIN. Network Neutrality: Modeling and Challenges and its Impact on Clouds: Keynote Talk, in "GECON 2017 - 14th International Conference on Economics of Grids, Clouds, Systems & Services", Biarritz, France, September 2017, https://hal.inria.fr/hal-01634382.
- [60] L. YALA, P. FRANGOUDIS, G. LUCARELLI, A. KSENTINI. Balancing between cost and availability for CD-NaaS resource placement, in "IEEE Global Communications Conference (GLOBECOM 2017)", Singapore, Singapore, December 2017, https://hal.inria.fr/hal-01590885.

### **National Conferences with Proceeding**

[61] N. RIVETTI, E. ANCEAUME, Y. BUSNEL, L. QUERZONI, B. SERICOLA. Ordonnancement dynamique pour un équilibrage de charge quasi-optimal dans les systèmes de traitement de flux, in "ALGOTEL 2017 - 19èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Quiberon, France, May 2017, https://hal.archives-ouvertes.fr/hal-01519432.

### **Conferences without Proceedings**

[62] Y. BUSNEL, N. RIVEEI, A. GAL.FlinkMan : Anomaly Detection in Manufacturing Equipment with Apache Flink : Grand Challenge, in "DEBS '17 - 11th ACM International Conference on Distributed and Eventbased Systems", Barcelone, Spain, June 2017, p. 274-279 [DOI : 10.1145/3093742.3095099], https://hal. archives-ouvertes.fr/hal-01644417.

### **Scientific Popularization**

[63] B. TUFFIN. La simulation de Monte-Carlo, in "Interstices", April 2017, https://hal.inria.fr/hal-01533686.

### **Other Publications**

- [64] I. ALAWE, Y. HADJADJ-AOUL, P. BERTIN, A. KSENTINI, D. DARCHE.RNN-based traffic prediction for pro-active scaling of the AMF, November 2017, 1, SDN Day 2017 - Software Defined Networks Day, Poster, https://hal.inria.fr/hal-01657683.
- [65] C. HARDY, E. LE MERRER, G. RUBINO, B. SERICOLA. Evaluation of Random Neural Layers in Deep Neural Networks, December 2017, 1, NIPS 2017 - workshop Deep Learning: Bridging Theory and Practice, Poster, https://hal.inria.fr/hal-01657644.
- [66] C. HARDY, E. LE MERRER, B. SERICOLA. Distributed deep learning on edge-devices in the Parameter Server Model, September 2017, 1, Workshop on Decentralized Machine Learning, Optimization and Privacy, Poster, https://hal.inria.fr/hal-01651145.
- [67] C. HARDY, E. LE MERRER, B. SERICOLA. *Distributed deep learning on edge-devices: feasibility via adaptive compression*, October 2017, working paper or preprint, https://hal.inria.fr/hal-01622580.
- [68] P. L'ECUYER, P. WAMBERGUE, E. BOURCERET. Spectral Analysis of the MIXMAX Random Number Generators, 2017, working paper or preprint, https://hal.inria.fr/hal-01634350.
- [69] P. MAILLÉ, B. TUFFIN. Analysis of Sponsored Data in the Case of Competing Wireless Service Providers, 2017, working paper or preprint, https://hal.inria.fr/hal-01670139.
- [70] P. MAILLÉ, B. TUFFIN. La fin des frais d'itinérance dans l'Union européenne en 2017, est-ce la panacée ?, 2017, working paper or preprint, https://hal.inria.fr/hal-01671481.
- [71] Y. MOCQUARD, B. SERICOLA, E. ANCEAUME. *Probabilistic Analysis of Rumor Spreading Time*, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01652777.

### **References in notes**

- [72] G. RUBINO, B. TUFFIN (editors). *Rare Event Simulation using Monte Carlo Methods*, John Wiley & Sons, 2009.
- [73] W. J. ANDERSON. Continuous-Time Markov Chains, Springer, 1991.

# **Project-Team DIVERSE**

# **Diversity-centric Software Engineering**

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: Institut national des sciences appliquées de Rennes Université Rennes 1

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Distributed programming and Software engineering

## **Table of contents**

2. Overall Objectives223. Research Program223.1. Scientific background223.1.1. Model-driven engineering223.1.2. Variability modeling223.1.3. Component-based software development223.1.4. Validation and verification233.1.5. Empirical software engineering233.2.1. Software Language Engineering233.2.1. Challenges233.2.2. Variability Modeling and Engineering233.2.2. Variability Modeling and Engineering233.2.3. Heterogeneous and dynamic software architectures233.2.3. Heterogeneous and dynamic software architectures233.2.4. Diverse implementations for resilience233.2.4. Diverse implementations for resilience233.2.4. Challenges233.2.4. Scientific objectives233.2.4. Diverse implementations for resilience233.2.4. Diverse implementations for resilience233.2.4. Challenges233.2.4. Challenges233.2.4. Diverse implementations for resilience233.2.4. Diverse implementations for resilience233.2.4. Challenges233.2.4. Diverse implementations for resilience233.2.4. Diverse implementations for resilience233.2.4. Scientific objectives233.2.4. Scientific objectives233.2.4. Scientific objectives233.2.4. Scientific objectives233.2.4. New Software and Platforms23	1.	Personnel	225		
3. Research Program 22   3.1. Scientific background 22   3.1. Model-driven engineering 22   3.1.1. Model-driven engineering 22   3.1.2. Variability modeling 22   3.1.3. Component-based software development 22   3.1.4. Validation and verification 23   3.1.5. Empirical software engineering 23   3.2. Research axis 23   3.2.1. Software Language Engineering 23   3.2.1.1. Challenges 23   3.2.2.2. Variability Modeling and Engineering 23   3.2.2.3. Cientific objectives 23   3.2.4. Challenges 23   3.2.5. Scientific objectives 23   3.2.6. Challenges 23   3.2.7. Scientific objectives 23   3.2.8. Heterogeneous and dynamic software architectures 23   3.2.3. Heterogeneous and dynamic software architectures 23   3.2.4. Diverse implementations for resilience 23   3.2.4.1. Challenges 23   3.2.4.2. Scientific objectives 23   3.2.4.1. Challenges 23   3.2.4.2. Scientific objectives 23   3.2.4.2	2.	Overall Objectives	226		
3.1. Scientific background223.1.1. Model-driven engineering223.1.2. Variability modeling223.1.3. Component-based software development223.1.4. Validation and verification233.1.5. Empirical software engineering233.2. Research axis233.2.1. Software Language Engineering233.2.1.2. Scientific objectives233.2.2. Variability Modeling and Engineering233.2.2. Variability Modeling and Engineering233.2.3. Heterogeneous and dynamic software architectures233.2.3. Lettific objectives233.2.4. Diverse implementations for resilience233.2.4. Diverse implementations for resilience233.2.4. Scientific objectives233.2.4. Challenges233.2.4. Scientific objectives233.2.5. Scientific objectives233.2.6. Scientific objectives233.2.7. Scientific objectives233.2.8. Scientific objectives233.2.9. Scientific objectives233.2.4. Diverse implementations for resilience233.2.4.2. Scientific objectives233.2.4.2. Scientific objectives233.2.4.2. Scientific objectives233.2.4.3. Challenges233.2.4.4. Challenges233.2.4.7. Scientific objectives233.2.4.8. Scientific objectives233.2.4.9. Scientific objectives233.2.4.1. Challenges233.2.4.2. Scientific objectives	3.	Research Program			
3.1.1.Model-driven engineering223.1.2.Variability modeling223.1.3.Component-based software development223.1.4.Validation and verification233.1.5.Empirical software engineering233.1.5.Empirical software engineering233.2.Research axis233.2.1.Software Language Engineering233.2.1.Challenges233.2.1.Challenges233.2.2.Variability Modeling and Engineering233.2.2.1.Challenges233.2.2.2.Scientific objectives233.2.2.3.Heterogeneous and dynamic software architectures233.2.3.1.Challenges233.2.3.2.Scientific objectives233.2.4.1.Challenges233.2.4.2.Scientific objectives233.2.4.3.Cientific objectives233.2.4.4.Challenges233.2.4.5.Scientific objectives233.2.4.6.Scientific objectives233.2.4.7.Scientific objectives233.2.4.8.Scientific objectives233.2.4.9.Scientific objectives233.2.4.1.Challenges233.2.4.2.Scientific objectives233.2.4.3.Scientific objectives233.2.4.4.Scientific objectives233.2.4.5.Scientific objectives233.2.4.6.Scientific objectives<		3.1. Scientific background	227		
3.1.2. Variability modeling223.1.3. Component-based software development223.1.4. Validation and verification233.1.5. Empirical software engineering233.2. Research axis233.2.1. Software Language Engineering233.2.1.1. Challenges233.2.1.2. Scientific objectives233.2.2. Variability Modeling and Engineering233.2.2.1. Challenges233.2.2.2. Scientific objectives233.2.3.1. Challenges233.2.3.2. Scientific objectives233.2.3.3. Heterogeneous and dynamic software architectures233.2.3.4. Diverse implementations for resilience233.2.4.1. Challenges233.2.4.2. Scientific objectives233.2.4.2. Scientific objectives233.2.4.3. Highlights of the Year234. New Software and Platforms235. Highlights of the Year233. Software and Platforms23		3.1.1. Model-driven engineering	227		
3.1.3.Component-based software development223.1.4.Validation and verification233.1.5.Empirical software engineering233.2.Research axis233.2.1.Software Language Engineering233.2.1.1.Challenges233.2.1.2.Scientific objectives233.2.2.Variability Modeling and Engineering233.2.2.1.Challenges233.2.2.2.Scientific objectives233.2.2.3.Heterogeneous and dynamic software architectures233.2.3.1.Challenges233.2.3.2.Scientific objectives233.2.3.2.Scientific objectives233.2.4.1.Challenges233.2.4.2.Scientific objectives233.2.4.2.Scientific objectives233.2.4.2.Scientific objectives233.2.4.3.Highlights of the Year234.Application Domains235.Highlights of the Year233.New Software and Platforms23		3.1.2. Variability modeling	228		
3.1.4. Validation and verification233.1.5. Empirical software engineering233.2. Research axis233.2.1. Software Language Engineering233.2.1.1. Challenges233.2.1.2. Scientific objectives233.2.2. Variability Modeling and Engineering233.2.2.1. Challenges233.2.2.2. Scientific objectives233.2.2.3. Heterogeneous and dynamic software architectures233.2.3.1. Challenges233.2.3.2. Scientific objectives233.2.3.3. Heterogeneous and dynamic software architectures233.2.3.4. Diverse implementations for resilience233.2.4.1. Challenges233.2.4.2. Scientific objectives233.2.4.2. Scientific objectives233.2.4.3. Highlights of the Year234. Application Domains235. Highlights of the Year236. New Software and Platforms23		3.1.3. Component-based software development			
3.1.5. Empirical software engineering233.2. Research axis233.2.1. Software Language Engineering233.2.1.1. Challenges233.2.1.2. Scientific objectives233.2.2. Variability Modeling and Engineering233.2.2.1. Challenges233.2.2.2. Scientific objectives233.2.3. Heterogeneous and dynamic software architectures233.2.3.1. Challenges233.2.3.2. Scientific objectives233.2.3.3. Heterogeneous and dynamic software architectures233.2.4.1. Challenges233.2.4.2. Scientific objectives233.2.4.1. Challenges233.2.4.2. Scientific objectives233.2.4.2. Scientific objectives233.2.4.3. Heterogeneous and dynamic software architectures233.2.4.4. Diverse implementations for resilience233.2.4.5. Scientific objectives233.2.4.6. Scientific objectives233.2.4.7. Scientific objectives233.2.4.8. Scientific objectives233.2.4.9. Scientific objectives233.2.4.2. Scientific objectives233.3.4.3. Scientific objectives233.3.4.4. Scientific objectives233.3.4.5. Scientific objectives233.3.4.6. Scientific objectives233.3.4.7. Scientific objectives233.3.4.8. Scientific objectives233.3.4.9. Scientific objectives233.3.4.1. Scientific objectives233.3.4.2. Scientific objectiv		3.1.4. Validation and verification	230		
3.2. Research axis233.2.1. Software Language Engineering233.2.1.1. Challenges233.2.1.2. Scientific objectives233.2.2. Variability Modeling and Engineering233.2.2.1. Challenges233.2.2.2. Scientific objectives233.2.3. Heterogeneous and dynamic software architectures233.2.3.1. Challenges233.2.3.2. Scientific objectives233.2.3.3. Heterogeneous and dynamic software architectures233.2.3.4. Diverse implementations for resilience233.2.4.1. Challenges233.2.4.2. Scientific objectives233.2.4.2. Scientific objectives233.2.4.3. Challenges233.2.4.4. Diverse implementations for resilience233.2.4.5. Scientific objectives233.2.4.6. Scientific objectives233.2.4.7. Scientific objectives233.2.4.8. Scientific objectives233.2.4.9. Scientific objectives233.2.4.2. Scientific objectives233.2.4.2. Scientific objectives233.3.4.2. Scientific objectives233.3.4.3. Scientific objectives233.3.4.3. Scientific objectives233.3.4.4.3. Scientific objectives233.3.4.5. Scientific objectives233.3.4.7. Scientific objectives233.3.4.7. Scientific objectives233.3.4.7. Scientific objectives233.3.4.7. Scientific objectives233.3.4.7. Scientific objectives23 </th <th></th> <th>3.1.5. Empirical software engineering</th> <th>231</th>		3.1.5. Empirical software engineering	231		
3.2.1.Software Language Engineering233.2.1.1.Challenges233.2.1.2.Scientific objectives233.2.2.Variability Modeling and Engineering233.2.2.1.Challenges233.2.2.2.Scientific objectives233.2.3.1.Challenges233.2.3.1.Challenges233.2.3.2.Scientific objectives233.2.3.3.1.Challenges233.2.4.1.Challenges233.2.4.2.Scientific objectives233.2.4.3.Scientific objectives233.2.4.4.Challenges233.2.4.5.Scientific objectives233.2.4.6.Scientific objectives233.2.4.7.Scientific objectives233.2.4.8.Scientific objectives233.2.4.9.Scientific objectives233.2.4.1.Challenges233.2.4.2.Scientific objectives233.2.4.3.Scientific objectives233.2.4.4.3.Scientific objectives233.2.4.4.4.Scientific objectives233.3.4.5.Scientific objectives233.3.6.6.New Software and Platforms233.3.7Scientific objectives333.3.7Scientific objectives333.3.7Scientific objectives333.3.7Scientific objectives333.3.7Scientific objectives333.3.7Scientific o		3.2. Research axis			
3.2.1.1.Challenges233.2.1.2.Scientific objectives233.2.2.1.Challenges233.2.2.2.Scientific objectives233.2.2.2.Scientific objectives233.2.3.1.Challenges233.2.3.2.Scientific objectives233.2.3.2.Scientific objectives233.2.3.1.Challenges233.2.3.2.Scientific objectives233.2.4.1.Challenges233.2.4.2.Scientific objectives233.2.4.2.Scientific objectives233.2.4.2.Scientific objectives233.2.4.3.Challenges233.2.4.4.Scientific objectives233.2.4.5.Scientific objectives233.2.4.6.Scientific objectives233.2.4.7.Scientific objectives233.2.4.8.Scientific objectives233.2.4.1.Challenges233.2.4.2.Scientific objectives233.2.4.3.Scientific objectives233.2.4.4.Scientific objectives233.2.4.2.Scientific objectives233.3.3.Scientific objectives233.3.3.Scientific objectives233.3.3.Scientific objectives233.3.3.Scientific objectives233.3.3.Scientific objectives233.3.3.Scientific objectives233.3.3.Scientific objectives<		3.2.1. Software Language Engineering	232		
3.2.1.2.Scientific objectives233.2.2.Variability Modeling and Engineering233.2.2.1.Challenges233.2.2.2.Scientific objectives233.2.3.1.Challenges233.2.3.2.Scientific objectives233.2.3.2.Scientific objectives233.2.3.2.Scientific objectives233.2.4.1.Challenges233.2.4.2.Scientific objectives233.2.4.2.Scientific objectives233.2.4.2.Scientific objectives233.2.4.2.Scientific objectives233.2.4.2.Scientific objectives233.2.4.2.Scientific objectives233.2.4.2.Scientific objectives233.2.4.3.Scientific objectives233.2.4.4.Scientific objectives233.2.4.2.Scientific objectives233.2.4.2.Scientific objectives233.2.4.2.Scientific objectives233.2.4.2.Scientific objectives233.2.4.3.Scientific objectives233.2.4.3.Scientific objectives233.3.4.4.3.Scientific objectives233.3.4.4.4.Scientific objectives233.3.5.Scientific objectives233.3.6.Scientific objectives233.3.7Scientific objectives233.3.7Scientific objectives233.3.7Scientific objectives23		3.2.1.1. Challenges	232		
3.2.2. Variability Modeling and Engineering233.2.2.1. Challenges233.2.2.2. Scientific objectives233.2.3. Heterogeneous and dynamic software architectures233.2.3.1. Challenges233.2.3.2. Scientific objectives233.2.4. Diverse implementations for resilience233.2.4.1. Challenges233.2.4.2. Scientific objectives233.2.4.2. Scientific objectives234. Application Domains235. Highlights of the Year236. New Software and Platforms23		3.2.1.2. Scientific objectives	232		
3.2.2.1.Challenges23.3.2.2.2.Scientific objectives23.3.2.3.1.Challenges23.3.2.3.2.Scientific objectives23.3.2.3.2.Scientific objectives23.3.2.4.1.Challenges23.3.2.4.2.Scientific objectives23.3.2.4.2.Scientific objectives23.3.2.4.3.Challenges23.3.2.4.4.3.Scientific objectives23.3.2.4.4.4.3.Scientific objectives23.3.2.4.5.Scientific objectives23.3.2.4.6.Scientific objectives23.3.2.4.7.Scientific objectives23.3.2.4.8.Scientific objectives23.3.2.4.9.Scientific objectives23.3.2.4.2.Scientific objectives23.3.2.4.3.Scientific objectives23.3.2.4.3.Scientific objectives23.3.2.4.3.Scientific objectives23.3.2.4.3.Scientific objectives23.3.2.4.3.Scientific objectives23.3.2.4.3.Scientific objectives23.3.3.3.Scientific objectives23.3.3.3.Scientific objectives23.3.3.3.Scientific objectives23.3.3.3.Scientific objectives23.3.3.3.Scientific objectives23.3.3.3.Scientific objectives23.3.3.3.Scientific objectives23.3.3.3.Scientific objectives<		3.2.2. Variability Modeling and Engineering	233		
3.2.2.2. Scientific objectives 23   3.2.3. Heterogeneous and dynamic software architectures 23   3.2.3.1. Challenges 23   3.2.3.2. Scientific objectives 23   3.2.4. Diverse implementations for resilience 23   3.2.4.1. Challenges 23   3.2.4.2. Scientific objectives 23   3.2.4.2. Scientific objectives 23   4. Application Domains 23   5. Highlights of the Year 23   6. New Software and Platforms 23		3.2.2.1. Challenges	233		
3.2.3. Heterogeneous and dynamic software architectures 23   3.2.3.1. Challenges 23   3.2.3.2. Scientific objectives 23   3.2.4. Diverse implementations for resilience 23   3.2.4.1. Challenges 23   3.2.4.2. Scientific objectives 23   3.2.4.2. Scientific objectives 23   3.2.4.3. Challenges 23   3.2.4.4. Challenges 23   3.2.4.5. Scientific objectives 23   3.2.4.6. Scientific objectives 23   3.2.4.7. Scientific objectives 23   3.2.4.8. Scientific objectives 23   4. Application Domains 23   5. Highlights of the Year 23   6. New Software and Platforms 23		3.2.2.2. Scientific objectives	233		
3.2.3.1. Challenges 23   3.2.3.2. Scientific objectives 23   3.2.4. Diverse implementations for resilience 23   3.2.4.1. Challenges 23   3.2.4.2. Scientific objectives 23   4. Application Domains 23   5. Highlights of the Year 23   6. New Software and Platforms 23		3.2.3. Heterogeneous and dynamic software architectures	234		
3.2.3.2. Scientific objectives 23   3.2.4. Diverse implementations for resilience 23   3.2.4.1. Challenges 23   3.2.4.2. Scientific objectives 23   4. Application Domains 23   5. Highlights of the Year 23   6. New Software and Platforms 23		3.2.3.1. Challenges	234		
3.2.4. Diverse implementations for resilience 23.   3.2.4.1. Challenges 23.   3.2.4.2. Scientific objectives 23.   4. Application Domains 23.   5. Highlights of the Year 23.   6. New Software and Platforms 23.		3.2.3.2. Scientific objectives	234		
3.2.4.1. Challenges 23.   3.2.4.2. Scientific objectives 23.   4. Application Domains 23.   5. Highlights of the Year 23.   6. New Software and Platforms 23.		3.2.4. Diverse implementations for resilience	235		
3.2.4.2. Scientific objectives 23   4. Application Domains 23   5. Highlights of the Year 23   6. New Software and Platforms 23		3.2.4.1. Challenges	235		
4. Application Domains 23   5. Highlights of the Year 23   6. New Software and Platforms 23		3.2.4.2. Scientific objectives	236		
5. Highlights of the Year 23'   6. New Software and Platforms 23'	4.	Application Domains	236		
6. New Software and Platforms	5.	Highlights of the Year	237		
	6.	New Software and Platforms	237		
6.1. amiunique 23		6.1. amiunique	237		
6.2. FAMILIAK 23		6.2. FAMILIAR	238		
6.3. GEMOC Studio 23		6.3. GEMOC Studio	238		
6.4. Kevoree 23		239			
6.5. Melange 24		6.5. Melange	240		
7 New Degulta	7	Now Desults	240		
7. New Results	/.	7.1 Besults on Variability modeling and management	241		
7.1.1. Veriability and testing		7.1. Veriability and testing	241		
7.1.2 Variability and teaching 24		7.1.2. Variability and teaching	241		
7.1.2. Variability and constraint colving 24		7.1.2. Variability and constraint solving	241		
7.1.4. Variability and machine learning (performance specialization of variability intensive)	7.1.3. variability and machine learning (performance specialization of variability)				
systeme)		systeme)	242		
7.1.5 Variability and machine learning (learning contextual variability models) 24		7.1.5 Variability and machine learning (learning contextual variability models)	242		
7.2 Results on Software Language Engineering		7.2 Results on Software Language Engineering	242		
7.2.1 On Language Interfaces 24		7.2.1 On Language Interfaces	242		
7.2.2. Revisiting Visitors for Modular Extension of Executable DSMLs 24		7.2.2. Revisiting Visitors for Modular Extension of Executable DSMLs	243		
7.2.3. Advanced and efficient execution trace management for executable domain-specific		ecific			
modeling languages 24		modeling languages	243		
7.2.4. Omniscient Debugging for Executable DSLs 24		7.2.4. Omniscient Debugging for Executable DSLs	243		
7.2.5. Reverse Engineering Language Product Lines from Existing DSL Variants 24		7.2.5. Reverse Engineering Language Product Lines from Existing DSL Variants	243		
7.2.6. Software Language Engineering for Virtual Reality Software Development 24		7.2.6. Software Language Engineering for Virtual Reality Software Development	244		
7.2.7. Create and Play your Pac-Man Game with the GEMOC Studio 24		7.2.7. Create and Play your Pac-Man Game with the GEMOC Studio	244		
7.3. Results on Heterogeneous and dynamic software architectures 24		7.3. Results on Heterogeneous and dynamic software architectures	244		

	7.3.1. Verifying the configuration of Virtualized Network Functions in Software I	Defined				
	Networks					
	7.3.2. Identity Negotiation at Runtime	245				
	7.3.3. Raising Time Awareness in Model-Driven Engineering	245				
	7.3.4. Collaborations					
	7.4. Results on Diverse Implementations for Resilience					
	7.4.1. Software diversification	246				
	7.4.2. Software testing	247				
8.	Bilateral Contracts and Grants with Industry	247				
	8.1.1. GLOSE	247				
	8.1.2. One Shot Software (OSS)	247				
9.	Partnerships and Cooperations	247				
	9.1. National Initiatives	247				
	9.1.1. ANR	247				
	9.1.1.1. SOPRANO	247				
	9.1.1.2. VaryVary ANR JCJC	248				
	9.1.1.3. CLARITY	248				
	9.1.1.4. Occiware	249				
	9.1.2. DGA	249				
	9.1.3. Cominiabs	249				
	9.2. European Initiatives	249				
	9.2.1. FP/& H2020 Projects	249				
	9.2.1.1. FP/ STREP HEADS	249				
	9.2.1.2. H2020 ICT-10-2016 STAMP	250				
	9.2.2. Collaborations with Major European Organizations	250				
	9.3. International Initiatives	250				
	9.3.1. Infla Associate Teams Not Involved in an Infla International Labs	250				
	9.3.2. Infla International Partners	251				
	9.5.5. Participation in Other International Programs	251				
	9.5.4. International Initiative GEMOC					
	9.4. International Research Visitors	252				
	9.4.2. Visits to International Teams	252				
10	9.4.2. Visits to international realits	252				
10.	10.1 Promoting Scientific Activities	252				
	10.1.1 Scientific Events Organisation	252				
	10.1.1.1 General Chair Scientific Chair	252				
	10.1.1.2 Member of the Organizing Committees	252				
	10.1.2. Scientific Events Selection	252				
	10.1.2.1 Member of the Conference Program Committees	252				
	10.1.2.2. Reviewer	253				
	10.1.3. Journal	253				
	10.1.3.1. Member of the Editorial Boards	253				
	10.1.3.2. Reviewer - Reviewing Activities	254				
	10.1.4. Invited Talks	2.54				
	10.1.5. Leadership within the Scientific Community	254				
	10.1.6. Scientific Expertise	254				
	10.1.7. Research Administration	254				
	10.2. Teaching - Supervision - Juries	255				
	10.2.1. Teaching	255				
	10.2.2. Supervision	255				

	10.2.3. Jurie	S	256
	10.2.3.1.	Benoit Baudry	256
	10.2.3.2.	Olivier Barais	256
	10.2.3.3.	Benoit Combemale	256
10.3. Popularization			
11.	Bibliography		256

### **Project-Team DIVERSE**

*Creation of the Team: 2014 January 01, updated into Project-Team: 2014 July 01* **Keywords:** 

### **Computer Science and Digital Science:**

A1.2.1. - Dynamic reconfiguration

A2.1.2. - Object-oriented programming

A2.1.10. - Domain-specific languages

A2.5. - Software engineering

A2.5.1. - Software Architecture & Design

A2.5.2. - Component-based Design

A2.5.3. - Empirical Software Engineering

A2.5.4. - Software Maintenance & Evolution

A2.6.2. - Middleware

A4.4. - Security of equipment and software

A4.8. - Privacy-enhancing technologies

### **Other Research Topics and Application Domains:**

B3.1. - Sustainable development

B3.1.1. - Resource management

B6.1. - Software industry

B6.1.1. - Software engineering

B6.1.2. - Software evolution, maintenance

B6.4. - Internet of things

B6.5. - Information systems

B6.6. - Embedded systems

B8.1.2. - Sensor networks for smart buildings

- B9.4.1. Computer science
- B9.8. Privacy

## 1. Personnel

### **Research Scientist**

Benoit Baudry [Team leader, Inria, Researcher, until Aug 2017, HDR]

### **Faculty Members**

Mathieu Acher [Univ de Rennes I, Associate Professor] Jean-Marc Jezequel [Univ de Rennes I, Professor, HDR] Olivier Barais [Univ de Rennes I, Professor, HDR] Arnaud Blouin [INSA Rennes, Associate Professor] Johann Bourcier [Univ de Rennes I, Associate Professor] Benoit Combemale [Univ de Rennes I, Associate Professor, until Aug 2017, HDR] Noel Plouzeau [Univ de Rennes I, Associate Professor]

### **External Collaborators**

Gurvan Le Guernic [DGA]

Benoit Combemale [Univ Toulouse, from Sep 2017, HDR]

### **Technical Staff**

Amine Benelallam [Univ de Rennes I] Fabien Coulon [Univ de Rennes I] Caroline Landry [Inria] Pierre Laperdrix [Inria, from Oct 2017] Tanja Mayerhofer [Univ de Rennes I, from Mar 2017 until May 2017] Maxime Tricoire [Inria] Andreas Wortmann [Univ de Rennes I, from Feb 2017 until May 2017]

### **PhD Students**

Mohamed Boussaa [Univ de Rennes I, Researcher, until Aug 2017] Kevin Corre [Orange Labs] Jean-Emile Dartois [Institut de recherche technologique B-com] Alejandro Gomez Boix [Inria] Nicolas Harrand [Inria] Pierre Laperdrix [Univ de Rennes I, until Sep 2017] Gwendal Le Moulec [INSA Rennes] Manuel Leduc [Univ de Rennes I] Dorian Leroy [Université étrangère] Ludovic Mouline [Gouvernement étranger] Youssou Ndiaye [Orange Labs] Johan Pelay [Institut de recherche technologique B-com] Quentin Plazar [Inria] Marcelino Rodriguez Cancio [Univ de Rennes I] Paul Temple [Univ de Rennes I] Oscar Luis Vera Perez [Inria]

### Visiting Scientist

Erwan Bousse [Université étrangère, from Sep 2017]

Administrative Assistant Tifenn Donguy [CNRS]

## 2. Overall Objectives

### 2.1. Overall objectives

DIVERSE's research agenda is in the area of software engineering. In this broad domain we develop models, methodologies and theories to address the challenges raised by the emergence of several forms of diversity in the design, deployment and evolution of software-intensive systems. The emergence of software diversity is an essential phenomenon in all application domains that we investigate with our industrial partners. These application domains range from complex systems such as systems of systems (in collaboration with Thales and DGA) and Instrumentation and Control (with EDF) to pervasive combinations of Internet of Things and Internet of Services (with TellU and Software AG) and tactical information systems (with the firefighter department). Even if today these systems are apparently radically different, we envision a strong convergence of the scientific principles underpinning their construction and validation towards **flexible and open yet dependable systems**. In particular, we see that the required flexibility and openness raise challenges for the software layer of these systems that must deal with four dimensions of diversity: the **diversity of languages** used by the stakeholders involved in the construction of these systems; the **diversity of features** required by the different customers; the **diversity of runtime environments** in which software has to run and adapt; the **diversity of implementations** that are necessary for resilience through redundancy.

In this context, the major software engineering challenge consists in handling **diversity** from variability in requirements and design to heterogeneous and dynamic execution environments. In particular this requires considering that the software system must adapt, in unpredictable ways, to changes in the requirements and environment. Conversely, explicitly handling of diversity is a great opportunity to allow software to spontaneously explore alternative design solutions. Concretely, we want to provide software engineers with the ability:

- to characterize an 'envelope' of possible variations
- to compose 'envelopes' (to discover new macro envelopes in an opportunistic manner)
- to dynamically synthesize software inside a given envelop

The major scientific objective that we must achieve to provide such mechanisms for software engineering is synthesized below

## Scientific objective for DIVERSE: Automatically compose and synthesize software diversity from design to runtime to address unpredictable evolutions of software-intensive systems

Software product lines and associated variability modeling formalisms represent an essential aspect of software diversity, which we already explored in the past and that represent a major foundation of DIVERSE's research agenda. However, DIVERSE also exploits other foundations to handle new forms of diversity: type theory and models of computation for the composition of languages; distributed algorithms and pervasive computation to handle the diversity of execution platforms; functional and qualitative randomized transformations to synthesize diversity for robust systems.

## 3. Research Program

### 3.1. Scientific background

### 3.1.1. Model-driven engineering

Model-Driven Engineering (MDE) aims at reducing the accidental complexity associated with developing complex software-intensive systems (e.g., use of abstractions of the problem space rather than abstractions of the solution space) [101]. It provides DIVERSE with solid foundations to specify, analyze and reason about the different forms of diversity that occur through the development lifecycle. A primary source of accidental complexity is the wide gap between the concepts used by domain experts and the low-level abstractions provided by general-purpose programming languages [72]. MDE approaches address this problem through modeling techniques that support separation of concerns and automated generation of major system artifacts from models (*e.g.*, test cases, implementations, deployment and configuration scripts). In MDE, a model describes an aspect of a system and is typically created or derived for specific development purposes [54]. Separation of concerns is supported through the use of different modeling languages, each providing constructs based on abstractions that are specific to an aspect of a system. MDE technologies also provide support for manipulating models, for example, support for querying, slicing, transforming, merging, and analyzing (including executing) models. Modeling languages are thus at the core of MDE, which participates to the development of a sound *Software Language Engineering*<sup>0</sup>, including an unified typing theory that integrate models as first class entities [104].

Incorporating domain-specific concepts and high-quality development experience into MDE technologies can significantly improve developer productivity and system quality. Since the late nineties, this realization has led to work on MDE language workbenches that support the development of domain-specific modeling languages (DSMLs) and associated tools (*e.g.*, model editors and code generators). A DSML provides a bridge between the field in which domain experts work and the implementation (programming) field. Domains in which DSMLs have been developed and used include, among others, automotive, avionics, and the emerging cyber-physical systems. A study performed by Hutchinson et al. [78] provides some indications that DSMLs can pave the way for wider industrial adoption of MDE.

<sup>&</sup>lt;sup>0</sup>See http://www.sleconf.org/

More recently, the emergence of new classes of systems that are complex and operate in heterogeneous and rapidly changing environments raises new challenges for the software engineering community. These systems must be adaptable, flexible, reconfigurable and, increasingly, self-managing. Such characteristics make systems more prone to failure when running and thus the development and study of appropriate mechanisms for continuous design and run-time validation and monitoring are needed. In the MDE community, research is focused primarily on using models at design, implementation, and deployment stages of development. This work has been highly productive, with several techniques now entering a commercialization phase. As software systems are becoming more and more dynamic, the use of model-driven techniques for validating and monitoring run-time behavior is extremely promising [86].

### 3.1.2. Variability modeling

While the basic vision underlying *Software Product Lines* (SPL) can probably be traced back to David Parnas seminal article [94] on the Design and Development of Program Families, it is only quite recently that SPLs are emerging as a paradigm shift towards modeling and developing software system families rather than individual systems [92]. SPL engineering embraces the ideas of mass customization and software reuse. It focuses on the means of efficiently producing and maintaining multiple related software products, exploiting what they have in common and managing what varies among them.

Several definitions of the *software product line* concept can be found in the research literature. Clements *et al.* define it as a set of software-intensive systems sharing a common, managed set of features that satisfy the specific needs of a particular market segment or mission and are developed from a common set of core assets in a prescribed way [91]. Bosch provides a different definition [60]: A SPL consists of a product line architecture and a set of reusable components designed for incorporation into the product line architecture. In addition, the PL consists of the software products developed using the mentioned reusable assets. In spite of the similarities, these definitions provide different perspectives of the concept: market-driven, as seen by Clements *et al.*, and *technology-oriented* for Bosch.

SPL engineering is a process focusing on capturing the *commonalities* (assumptions true for each family member) and *variability* (assumptions about how individual family members differ) between several software products [66]. Instead of describing a single software system, a SPL model describes a set of products in the same domain. This is accomplished by distinguishing between elements common to all SPL members, and those that may vary from one product to another. Reuse of core assets, which form the basis of the product line, is key to productivity and quality gains. These core assets extend beyond simple code reuse and may include the architecture, software components, domain models, requirements statements, documentation, test plans or test cases.

The SPL engineering process consists of two major steps:

- 1. **Domain Engineering**, or *development for reuse*, focuses on core assets development.
- 2. **Application Engineering**, or *development with reuse*, addresses the development of the final products using core assets and following customer requirements.

Central to both processes is the management of **variability** across the product line [74]. In common language use, the term *variability* refers to *the ability or the tendency to change*. Variability management is thus seen as the key feature that distinguishes SPL engineering from other software development approaches [61]. Variability management is thus growingly seen as the cornerstone of SPL development, covering the entire development life cycle, from requirements elicitation [106] to product derivation [111] to product testing [90], [89].

Halmans *et al.* [74] distinguish between *essential* and *technical* variability, especially at requirements level. Essential variability corresponds to the customer's viewpoint, defining what to implement, while technical variability relates to product family engineering, defining how to implement it. A classification based on the dimensions of variability is proposed by Pohl *et al.* [96]: beyond **variability in time** (existence of different versions of an artifact that are valid at different times) and **variability in space** (existence of an artifact in different shapes at the same time) Pohl *et al.* claim that variability is important to different stakeholders and thus has different levels of visibility: **external variability** is visible to the customers while **internal** 

**variability**, that of domain artifacts, is hidden from them. Other classification proposals come from Meekel *et al.* [84] (feature, hardware platform, performances and attributes variability) or Bass *et al.* [52] who discuss about variability at the architectural level.

Central to the modeling of variability is the notion of *feature*, originally defined by Kang *et al.* as: *a prominent* or distinctive user-visible aspect, quality or characteristic of a software system or systems [80]. Based on this notion of *feature*, they proposed to use a *feature model* to model the variability in a SPL. A feature model consists of a *feature diagram* and other associated information: constraints and dependency rules. Feature diagrams provide a graphical tree-like notation depicting the hierarchical organization of high level product functionalities represented as features. The root of the tree refers to the complete system and is progressively decomposed into more refined features (tree nodes). Relations between nodes (features) are materialized by decomposition edges and textual constraints. Variability can be expressed in several ways. Presence or absence of a feature from a product is modeled using mandatory or optional features. Features are graphically represented as rectangles while some graphical elements (e.g., unfilled circle) are used to describe the variability (e.g., a feature may be optional).

Features can be organized into *feature groups*. Boolean operators *exclusive alternative (XOR)*, *inclusive alternative (OR)* or *inclusive (AND)* are used to select one, several or all the features from a feature group. Dependencies between features can be modeled using *textual constraints: requires* (presence of a feature requires the presence of another), *mutex* (presence of a feature automatically excludes another). Feature attributes can be also used for modeling quantitative (e.g., numerical) information. Constraints over attributes and features can be specified as well.

Modeling variability allows an organization to capture and select which version of which variant of any particular aspect is wanted in the system [61]. To implement it cheaply, quickly and safely, redoing by hand the tedious weaving of every aspect is not an option: some form of automation is needed to leverage the modeling of variability [56], [68]. Model Driven Engineering (MDE) makes it possible to automate this weaving process [79]. This requires that models are no longer informal, and that the weaving process is itself described as a program (which is as a matter of facts an executable meta-model [87]) manipulating these models to produce for instance a detailed design that can ultimately be transformed to code, or to test suites [95], or other software artifacts.

### 3.1.3. Component-based software development

Component-based software development [105] aims at providing reliable software architectures with a low cost of design. Components are now used routinely in many domains of software system designs: distributed systems, user interaction, product lines, embedded systems, etc. With respect to more traditional software artifacts (e.g., object oriented architectures), modern component models have the following distinctive features [67]: description of requirements on services required from the other components; indirect connections between components thanks to ports and connectors constructs [82]; hierarchical definition of components (assemblies of components can define new component types); connectors supporting various communication semantics [64]; quantitative properties on the services [59].

In recent years component-based architectures have evolved from static designs to dynamic, adaptive designs (e.g., SOFA [64], Palladio [57], Frascati [88]). Processes for building a system using a statically designed architecture are made of the following sequential lifecycle stages: requirements, modeling, implementation, packaging, deployment, system launch, system execution, system shutdown and system removal. If for any reason after design time architectural changes are needed after system launch (e.g., because requirements changed, or the implementation platform has evolved, etc) then the design process must be reexecuted from scratch (unless the changes are limited to parameter adjustment in the components deployed).

Dynamic designs allow for *on the fly* redesign of a component based system. A process for dynamic adaptation is able to reapply the design phases while the system is up and running, without stopping it (this is different from stop/redeploy/start). This kind of process supports *chosen adaptation*, when changes are planned and realized to maintain a good fit between the needs that the system must support and the way it supports them [81]. Dynamic component-based designs rely on a component meta-model that supports complex life cycles

for components, connectors, service specification, etc. Advanced dynamic designs can also take platform changes into account at run-time, without human intervention, by adapting themselves [65], [108]. Platform changes and more generally environmental changes trigger *imposed adaptation*, when the system can no longer use its design to provide the services it must support. In order to support an eternal system [58], dynamic component based systems must separate architectural design and platform compatibility. This requires support for heterogeneity, since platform evolutions can be partial.

The Models@runtime paradigm denotes a model-driven approach aiming at taming the complexity of dynamic software systems. It basically pushes the idea of reflection one step further by considering the reflection layer as a real model "something simpler, safer or cheaper than reality to avoid the complexity, danger and irreversibility of reality [99]". In practice, component-based (and/or service-based) platforms offer reflection APIs that make it possible to introspect the system (which components and bindings are currently in place in the system) and dynamic adaptation (by applying CRUD operations on these components and bindings). While some of these platforms offer rollback mechanisms to recover after an erroneous adaptation, the idea of Models@runtime is to prevent the system from actually enacting an erroneous adaptation. In other words, the "model at run-time" is a reflection model that can be uncoupled (for reasoning, validation, simulation purposes) and automatically resynchronized.

Heterogeneity is a key challenge for modern component based system. Until recently, component based techniques were designed to address a specific domain, such as embedded software for command and control, or distributed Web based service oriented architectures. The emergence of the Internet of Things paradigm calls for a unified approach in component based design techniques. By implementing an efficient separation of concern between platform independent architecture management and platform dependent implementations, *Models@runtime* is now established as a key technique to support dynamic component based designs. It provides DIVERSE with an essential foundation to explore an adaptation envelop at run-time.

Search Based Software Engineering [76] has been applied to various software engineering problems in order to support software developers in their daily work. The goal is to automatically explore a set of alternatives and assess their relevance with respect to the considered problem. These techniques have been applied to craft software architecture exhibiting high quality of services properties [73]. Multi Objectives Search based techniques [70] deal with optimization problem containing several (possibly conflicting) dimensions to optimize. These techniques provide DIVERSE with the scientific foundations for reasoning and efficiently exploring an envelope of software configurations at run-time.

### 3.1.4. Validation and verification

Validation and verification (V&V) theories and techniques provide the means to assess the validity of a software system with respect to a specific correctness envelop. As such, they form an essential element of DIVERSE's scientific background. In particular, we focus on model-based V&V in order to leverage the different models that specify the envelop at different moments of the software development lifecycle.

Model-based testing consists in analyzing a formal model of a system (*e.g.*, activity diagrams, which capture high-level requirements about the system, statecharts, which capture the expected behavior of a software module, or a feature model, which describes all possible variants of the system) in order to generate test cases that will be executed against the system. Model-based testing [107] mainly relies on model analysis, constraint solving [69] and search-based reasoning [83]. DIVERSE leverages in particular the applications of model-based testing in the context of highly-configurable systems and [109] interactive systems [85] as well as recent advances based on diversity for test cases selection [77].

Nowadays, it is possible to simulate various kinds of models. Existing tools range from industrial tools such as Simulink, Rhapsody or Telelogic to academic approaches like Omega [93], or Xholon <sup>0</sup>. All these simulation environments operate on homogeneous environment models. However, to handle diversity in software systems, we also leverage recent advances in heterogeneous simulation. Ptolemy [63] proposes a common abstract syntax, which represents the description of the model structure. These elements can be

<sup>&</sup>lt;sup>0</sup>http://www.primordion.com/Xholon/

decorated using different directors that reflect the application of a specific model of computation on the model element. Metropolis [53] provides modeling elements amenable to semantically equivalent mathematical models. Metropolis offers a precise semantics flexible enough to support different models of computation. ModHel'X [75] studies the composition of multi-paradigm models relying on different models of computation.

Model-based testing and simulation are complemented by runtime fault-tolerance through the automatic generation of software variants that can run in parallel, to tackle the open nature of software-intensive systems. The foundations in this case are the seminal work about N-version programming [51], recovery blocks [97] and code randomization [55], which demonstrated the central role of diversity in software to ensure runtime resilience of complex systems. Such techniques rely on truly diverse software solutions in order to provide systems with the ability to react to events, which could not be predicted at design time and checked through testing or simulation.

### 3.1.5. Empirical software engineering

The rigorous, scientific evaluation of DIVERSE's contributions is an essential aspect of our research methodology. In addition to theoretical validation through formal analysis or complexity estimation, we also aim at applying state-of-the-art methodologies and principles of empirical software engineering. This approach encompasses a set of techniques for the sound validation contributions in the field of software engineering, ranging from statistically sound comparisons of techniques and large-scale data analysis to interviews and systematic literature reviews [102], [100]. Such methods have been used for example to understand the impact of new software development paradigms [62]. Experimental design and statistical tests represent another major aspect of empirical software engineering. Addressing large-scale software engineering problems often requires the application of heuristics, and it is important to understand their effects through sound statistical analyses [50].

### 3.2. Research axis

Figure 1 illustrates the four dimensions of software diversity, which form the core research axis of DIVERSE: the **diversity of languages** used by the stakeholders involved in the construction of these systems; the **diversity of features** required by the different customers; the **diversity of runtime environments** in which software has to run and adapt; the **diversity of implementations** that are necessary for resilience through redundancy. These four axis share and leverage the scientific and technological results developed in the area of model-driven engineering in the last decade. This means that all our research activities are founded on sound abstractions to reason about specific aspects of software systems, compose different perspectives and automatically generate parts of the system.



Figure 1. The four research axis of DIVERSE, which rely on a MDE scientific background

### 3.2.1. Software Language Engineering

The engineering of systems involves many different stakeholders, each with their own domain of expertise. Hence more and more organizations are adopting Domain Specific Modeling Languages (DSMLs) to allow domain experts to express solutions directly in terms of relevant domain concepts [101], [72]. This new trend raises new challenges about designing DSMLs, evolving a set of DSMLs and coordinating the use of multiple DSLs for both DSL designers and DSL users.

### 3.2.1.1. Challenges

**Reusability** of software artifacts is a central notion that has been thoroughly studied and used by both academics and industrials since the early days of software construction. Essentially, designing reusable artifacts allows the construction of large systems from smaller parts that have been separately developed and validated, thus reducing the development costs by capitalizing on previous engineering efforts. However, it is still hardly possible for language designers to design typical language artifacts (e.g. language constructs, grammars, editors or compilers) in a reusable way. The current state of the practice usually prevents the reusability of language artifacts from one language to another, consequently hindering the emergence of real engineering techniques around software languages. Conversely, concepts and mechanisms that enable artifacts reusability abound in the software engineering community.

**Variability** in modeling languages occur in the definition of the abstract and concrete syntax as well as in the specification of the language's semantics. The major challenges met when addressing the need for variability are: (i) set principles for modeling language units that support the modular specification of a modeling language; and (ii) design mechanisms to assemble these units in a complete language, according to the set of authorized variation points for the modeling language family.

A new generation of complex software-intensive systems (for example smart health support, smart grid, building energy management, and intelligent transportation systems) presents new opportunities for leveraging modeling languages. The development of these systems requires expertise in diverse domains. Consequently, different types of stakeholders (e.g., scientists, engineers and end-users) must work in a coordinated manner on various aspects of the system across multiple development phases. DSMLs can be used to support the work of domain experts who focus on a specific system aspect, but they can also provide the means for coordinating work across teams specializing in different aspects and across development phases. The support and integration of DSMLs leads to what we call **the globalization of modeling languages**, *i.e.* the use of multiple languages for the coordinated development of diverse aspects of a system. One can make an analogy with world globalization in which relationships are established between sovereign countries to regulate interactions (e.g., travel and commerce related interactions) while preserving each country's independent existence.

### 3.2.1.2. Scientific objectives

We address reuse and variability challenges through the investigation of the time-honored concepts of substitutability, inheritance and components, evaluate their relevance for language designers and provide tools and methods for their inclusion in software language engineering. We will develop novel techniques for the modular construction of language extensions with the support of model syntactical variability. From the semantics perspective, we investigate extension mechanisms for the specification of variability in operational semantics, focusing on static introduction and heterogeneous models of computation. The definition of variation points for the three aspects of the language definition provides the foundations for the novel concept Language Unit (LU) as well as suitable mechanisms to compose such units.

We explore the necessary breakthrough in software languages to support modeling and simulation of heterogeneous and open systems. This work relies on the specification of executable domain specific modeling languages (DSMLs) to formalize the various concerns of a software-intensive system, and of models of computation (MoCs) to explicitly model the concurrency, time and communication of such DSMLs. We develop a framework that integrates the necessary foundations and facilities for designing and implementing executable and concurrent domain-specific modeling languages. It also provides unique features to specify composition operators between (possibly heterogeneous) DSMLs. Such specifications are amenable to support the edition, execution, graphical animation and analysis of heterogeneous models. The objective is to provide both a significant improvement of MoCs and DSMLs design and implementation; and the simulation based validation and verification of complex systems.

We see an opportunity for the automatic diversification of programs' computation semantics, for example through the diversification of compilers or virtual machines. The main impact of this artificial diversity is to provide flexible computation and thus ease adaptation to different execution conditions. A combination of static and dynamic analysis could support the identification of what we call *plastic computation zones* in the code. We identify different categories of such zones: (i) areas in the code in which the order of computation can vary (e.g., the order in which a block of sequential statements is executed); (ii) areas that can be removed, keeping the essential functionality [103] (e.g., skip some loop iterations); (iii) areas that can replaced by alternative code (e.g., replace a try-catch by a return statement). Once we know which zones in the code can be randomized, it is necessary to modify the model of computation to leverage the computation plasticity. This consists in introducing variation points in the interpreter to reflect the diversity of models of computation. Then, the choice of a given variation is performed randomly at run-time.

### 3.2.2. Variability Modeling and Engineering

The systematic modeling of variability in software systems has emerged as an effective approach to document and reason about software evolutions and heterogeneity (*cf.* Section 3.1.2). Variability modeling characterizes an "envelope" of possible software variations. The industrial use of variability models and their relation to software artifact models require a complete engineering framework, including composition, decomposition, analysis, configuration and artifact derivation, refactoring, re-engineering, extraction, and testing. This framework can be used both to tame imposed diversity and to manage chosen diversity.

### 3.2.2.1. Challenges

A fundamental problem is that the **number of variants** can be exponential in the number of options (features). Already with 300 boolean configuration options, approximately  $10^{90}$  configurations exist – more than estimated count of atoms in the universe. Domains like automotive or operating systems have to manage more than 10000 options (e.g., Linux). Practitioners face the challenge of developing billions of variants. It is easy to forget a necessary constraint, leading to the synthesis of unsafe variants, or to under-approximate the capabilities of the software platform. Scalable modelling techniques are therefore crucial to specify and reason about a very large set of variants.

Model-driven development supports two ways to deal with the increasing number of concerns in complex systems: (1) multi-view modeling, *i.e.* when modeling each concern separately, and variability modeling. However, there is little support to combine both approaches consistently. Techniques to integrate both approaches will enable the construction of a consistent set of views and variation points in each view.

The design, construction and maintenance of software families have a major impact on **software testing**. Among the existing challenges, we can cite: the selection of test cases for a specific variant; the evolution of test suites with integration of new variants; the combinatorial explosion of the number of software configurations to be tested. Novel model-based techniques for test generation and test management in a software product line context are needed to overcome state-of-the-art limits we already observed in some projects.

#### 3.2.2.2. Scientific objectives

We aim at developing scalable techniques to automatically analyze variability models and their interactions with other views on the software intensive system (requirements, architecture, design). These techniques provide two major advancements in the state of the art: (1) an extension of the semantics of variability models in order to enable the definition of attributes (*e.g.*, cost, quality of service, effort) on features and to include these attributes in the reasoning; (2) an assessment of the consistent specification of variability models with respect to system views (since variability is orthogonal to system modeling, it is currently possible to specify the different models in ways that are semantically meaningless). The former aspect of analysis is tackled through constraint solving and finite-domain constraint programming, while the latter aspect is investigated through automatic search-based techniques (similar to genetic algorithms) for the exploration of the space of interaction between variability and view models.

We aim to develop procedures to reverse engineer dependencies and features' sets from existing software artefacts – be it source code, configuration files, spreadsheets (e.g., product comparison matrices) or requirements. We expect to scale up (e.g., for extracting a very large number of variation points) and guarantee some properties (e.g., soundness of configuration semantics, understandability of ontological semantics). For instance, when building complex software-intensive systems, textual requirements are captured in very large quantities of documents. In this context, adequate models to formalize the organization of requirements documents and automated techniques to support impact analysis (in case of changes in the requirements) have to be developed.

We aim at developing sound methods and tools to integrate variability management in model-based testing activities. In particular, we will leverage requirement models as an essential asset to establish formal relations between variation points and test models. These relations will form the basis for novel algorithms that drive the systematic selection of test configurations that satisfy well-defined test adequacy criteria as well as the generation of test cases for a specific product in the product line.

### 3.2.3. Heterogeneous and dynamic software architectures

Flexible yet dependable systems have to cope with heterogeneous hardware execution platforms ranging from smart sensors to huge computation infrastructures and data centers. Evolutions range from a mere change in the system configuration to a major architectural redesign, for instance to support addition of new features or a change in the platform architecture (new hardware is made available, a running system switches to low bandwidth wireless communication, a computation node battery is running low, etc.). In this context, we need to devise formalisms to reason about the impact of an evolution and about the transition from one configuration to another. It must be noted that this axis focuses on the use of models to drive the evolution from design time to run-time. Models will be used to (i) systematically define predictable configurations and variation points through which the system will evolve; (ii) develop behaviors necessary to handle unpredicted evolutions.

### 3.2.3.1. Challenges

The main challenge is to provide new homogeneous architectural modelling languages and efficient techniques that enable continuous software reconfiguration to react to changes. This work handles the challenges of handling the diversity of runtime infrastructures and managing the cooperation between different stakeholders. More specifically, the research developed in this axis targets the following dimensions of software diversity.

Platform architectural heterogeneity induces a first dimension of imposed diversity (type diversity). Platform reconfigurations driven by changing resources define another dimension of diversity (deployment diversity). To deal with these imposed diversity problems, we will rely on model based runtime support for adaptation, in the spirit of the dynamic distributed component framework developed by the Triskell team. Since the runtime environment composed of distributed, resource constrained hardware nodes cannot afford the overhead of traditional runtime adaptation techniques, we investigate the design of novel solutions relying on models@runtime and on specialized tiny virtual machines to offer resource provisioning and dynamic reconfigurations. In the next two years this research will be supported by the InfraJVM project.

Diversity can also be an asset to optimize software architecture. Architecture models must integrate multiple concerns in order to properly manage the deployment of software components over a physical platform. However, these concerns can contradict each other (*e.g.*, accuracy and energy). In this context, we investigate automatic solutions to explore the set of possible architecture models and to establish valid trade-offs between all concerns in case of changes.

#### 3.2.3.2. Scientific objectives

Automatic synthesis of optimal software architectures. Implementing a service over a distributed platform (*e.g.*, a pervasive system or a cloud platform) consists in deploying multiple software components over distributed computation nodes. We aim at designing search-based solutions to (i) assist the software architect in establishing a good initial architecture (that balances between different factors such as cost of the nodes, latency, fault tolerance) and to automatically update the architecture when the environment or the system itself change. The choice of search-based techniques is motivated by the very large number of possible software deployment architectures that can be investigated and that all provide different trade-offs between qualitative

factors. Another essential aspect that is supported by multi-objective search is to explore different architectural solutions that are not necessarily comparable. This is important when the qualitative factors are orthogonal to each other, such as security and usability for example.

Flexible software architecture for testing and data management. As the number of platforms on which software runs increases and different software versions coexist, the demand for testing environments also increases. For example, to test a software patch or upgrade, the number of testing environments is the product of the number of running environments the software supports and the number of coexisting versions of the software. Based on our first experiment on the synthesis of cloud environment using architectural models, our objective is to define a set of domain specific languages to catch the requirement and to design cloud environments for testing and data management of future internet systems from data centers to things. These languages will be interpreted to support dynamic synthesis and reconfiguration of a testing environment.

**Runtime support for heterogeneous environments.** Execution environments must provide a way to account or reserve resources for applications. However, current execution environments such as the Java Virtual Machine do not clearly define a notion of application: each framework has its own definition. For example, in OSGi, an application is a component, in JEE, an application is most of the time associated to a class loader, in the Multi-Tasking Virtual machine, an application is a process. The challenge consists in defining an execution environment that provides direct control over resources (CPU, Memory, Network I/O) independently from the definition of an application. We propose to define abstract resource containers to account and reserve resources on a distributed network of heterogeneous devices.

### 3.2.4. Diverse implementations for resilience

Open software-intensive systems have to evolve over their lifetime in response to changes in their environment. Yet, most verification techniques assume a closed environment or the ability to predict all changes. Dynamic changes and evolutions thus represent a major challenge for these techniques that aim at assessing the correctness and robustness of the system. On the one hand, DIVERSE will adapt V&V techniques to handle diversity imposed by the requirements and the execution environment, on the other hand we leverage diversity to increase the robustness of software in face of unpredicted situations. More specifically, we address the following V&V challenges.

### 3.2.4.1. Challenges

One major challenge to build flexible and open yet dependable systems is that current software engineering techniques require architects to foresee all possible situations the system will have to face. However, openness and flexibility also mean unpredictability: unpredictable bugs, attacks, environmental evolutions, etc. Current fault-tolerance [97] and security [71] techniques provide software systems with the capacity of detecting accidental and deliberate faults. However, existing solutions assume that the set of bugs or vulnerabilities in a system does not evolve. This assumption does not hold for open systems, thus it is essential to revisit fault-tolerance and security solutions to account for diverse and unpredictable faults.

Diversity is known to be a major asset for the robustness of large, open, and complex systems (*e.g.*, economical or ecological systems). Following this observation, the software engineering literature provides a rich set of work that choose to implement diversity in software systems in order to improve robustness to attacks or to changes in quality of service. These works range from N-version programming to obfuscation of data structures or control flow, to randomization of instruction sets. An essential remaining challenge is to support the automatic synthesis and evolution of software diversity in open software-intensive systems. There is an opportunity to further enhance these techniques in order to cope with a wider diversity of faults, by multiplying the levels of diversity in the different software layers that are found in software-intensive systems (system, libraries, frameworks, application). This increased diversity must be based on artificial program transformations and code synthesis, which increase the chances of exploring novel solutions, better fitted at one point in time. The biological analogy also indicates that diversity should emerge as a side-effect of evolution, to prevent over-specialization towards one kind of diversity.

#### 3.2.4.2. Scientific objectives

The main objective is to address one of the main limitations of N-version programming for fault-tolerant systems: the manual production and management of software diversity. Through automated injection of artificial diversity we aim at systematically increasing failure diversity and thus increasing the chances of early error detection at run-time. A fundamental assumption for this work is that software-intensive systems can be "good enough" [98], [110].

**Proactive program diversification.** We aim at establishing novel principles and techniques that favor the emergence of multiple forms of software diversity in software-intensive systems, in conjunction with the software adaptation mechanisms that leverage this diversity. The main expected outcome is a set of metadesign principles that maintain diversity in systems and the experimental demonstration of the effects of software diversity on the adaptive capacities of CASs. Higher levels of diversity in the system provide a pool of software solutions that can eventually be used to adapt to situations unforeseen at design time (bugs, crash, attacks, etc.). Principles of automated software diversification rely on the automated synthesis of variants in a software product line, as well as finer-grained program synthesis combining unsound transformations and genetic programming to explore the space of mutational robustness.

**Multi-tier software diversification.** We call multi-tier diversification the fact of diversifying several application software components simultaneously. The novelty of our proposal, with respect to the software diversity state of the art, is to diversify the application-level code (for example, diversify the business logics of the application), focusing on the technical layers found in web applications. The diversification of application software code is expected to provide a diversity of failures and vulnerabilities in web server deployment. Web server deployment usually adopts a form of the Reactor architecture pattern, for scalability purposes: multiple copies of the server software stack, called request handlers, are deployed behind a load balancer. This architecture is very favorable for diversification, since by using the multiplicity of request handlers running in a web server we can simultaneously deploy multiple combinations of diverse software components. Then, if one handler is hacked or crashes the others should still be able to process client requests.

## 4. Application Domains

### 4.1. From Embedded Systems to Service Oriented Architectures

From small embedded systems such as home automation products or automotive systems to medium sized systems such as medical equipment, office equipment, household appliances, smart phones; up to large Service Oriented Architectures (SOA), building a new application from scratch is no longer possible. Such applications reside in (group of) machines that are expected to run continuously for years without unrecoverable errors. Special care has then to be taken to design and validate embedded software, making the appropriate trade-off between various extra-functional properties such as reliability, timeliness, safety and security but also development and production cost, including resource usage of processor, memory, bandwidth, power, etc.

Leveraging ongoing advances in hardware, embedded software is playing an evermore crucial role in our society, bound to increase even more when embedded systems get interconnected to deliver ubiquitous SOA. For this reason, embedded software has been growing in size and complexity at an exponential rate for the past 20 years, pleading for a component based approach to embedded software development. There is a real need for flexible solutions allowing to deal at the same time with a wide range of needs (product lines modeling and methodologies for managing them), while preserving quality and reducing the time to market (such as derivation and validation tools).

We believe that building flexible, reliable and efficient embedded software will be achieved by reducing the gap between executable programs, their models, and the platform on which they execute, and by developing new composition mechanisms as well as transformation techniques with a sound formal basis for mapping between the different levels.

Reliability is an essential requirement in a context where a huge number of softwares (and sometimes several versions of the same program) may coexist in a large system. On one hand, software should be able to evolve very fast, as new features or services are frequently added to existing ones, but on the other hand, the occurrence of a fault in a system can be very costly, and time consuming. While we think that formal methods may help solving this kind of problems, we develop approaches where they are kept "behind the scene" in a global process taking into account constraints and objectives coming from user requirements.

Software testing is another aspect of reliable development. Testing activities mostly consist in trying to exhibit cases where a system implementation does not conform to its specifications. Whatever the efforts spent for development, this phase is of real importance to raise the confidence level in the fact that a system behaves properly in a complex environment. We also put a particular emphasis on on-line approaches, in which test and observation are dynamically computed during execution.

## 5. Highlights of the Year

### 5.1. Highlights of the Year

### 5.1.1. Awards

- Publications
  - Learning-Contextual Variability Models, IEEE Software
  - Correctness Attraction: A Study of Stability of Software Behavior Under Runtime Perturbation, Empirical Software Engineering

item Great positions for members (KTH, Univ Toulouse, Mc Gill, ...)

- Three new direct collaborations with Industrial partners: Orange, Nokia, Safran
- Great visibility for AmIUnique and several popularization actions
- Kermeta transfer to Obeo

## 6. New Software and Platforms

### 6.1. amiunique

**KEYWORDS:** Privacy - Browser fingerprinting

SCIENTIFIC DESCRIPTION: The amiunique web site has been deployed in the context of the DiverSE's research activities on browser fingerprinting and how software diversity can be leveraged in order to mitigate the impact of fingerprinting on the privacy of users. The construction of a dataset of genuine fingerprints is essential to understand in details how browser fingerprints can serve as unique identifiers and hence what should be modified in order to mitigate its impact privacy. This dataset also supports the large-scale investigation of the impact of web technology advances on fingerprinting. For example, we can analyze in details the impact of the HTML5 canvas element or the behavior of fingerprinting on mobile devices.

The whole source code of amiunique is open source and is distributed under the terms of the MIT license.

Similar sites: Panopticlick https://panopticlick.eff.org/ BrowserSpy http://browserspy.dk/ http://noc.to/ Main innovative features: canvas fingerprinting WebGL fingerprinting advanced JS features (platform, DNT, etc.)

Impact: The website has been showcased in several professional forums in 2014 and 2015 (Open World Forum 2014, FOSSA'14, FIC'15, ICT'15) and it has been visited by more than 100000 unique visitors in one year.

FUNCTIONAL DESCRIPTION: This web site aims at informing visitors about browser fingerprinting and possible tools to mitigate its effect, as well as at collecting data about the fingerprints that can be found on the web. It collects browser fingerprints with the explicit agreement of the users (they have to click on a button on the home page). Fingerprints are composed of 17 attributes, which include regular HTTP headers as well as the most recent state of the art techniques (canvas fingerprinting, WebGL information).

- Participants: Benoit Baudry and Pierre Laperdrix
- Partner: INSA Rennes
- Contact: Benoit Baudry
- URL: https://amiunique.org/

### 6.2. FAMILIAR

KEYWORDS: Software line product - Configators - Customisation

SCIENTIFIC DESCRIPTION: FAMILIAR (for FeAture Model scrIpt Language for manIpulation and Automatic Reasoning) is a language for importing, exporting, composing, decomposing, editing, configuring, computing "diffs", refactoring, reverse engineering, testing, and reasoning about (multiple) feature models. All these operations can be combined to realize complex variability management tasks. A comprehensive environment is proposed as well as integration facilities with the Java ecosystem.

FUNCTIONAL DESCRIPTION: Familiar is an environment for large-scale product customisation. From a model of product features (options, parameters, etc.), Familiar can automatically generate several million variants. These variants can take many forms: software, a graphical interface, a video sequence or even a manufactured product (3D printing). Familiar is particularly well suited for developing web configurators (for ordering customised products online), for providing online comparison tools and also for engineering any family of embedded or software-based products.

- Participants: Aymeric Hervieu, Benoit Baudry, Didier Vojtisek, Edward Mauricio Alferez Salinas, Guillaume Bécan, Joao Bosco Ferreira-Filho, Julien Richard-Foy, Mathieu Acher, Olivier Barais and Sana Ben Nasr
- Contact: Mathieu Acher
- URL: http://familiar-project.github.com

### 6.3. GEMOC Studio

KEYWORDS: DSL - Language workbench - Model debugging

SCIENTIFIC DESCRIPTION: The language workbench put together the following tools seamlessly integrated to the Eclipse Modeling Framework (EMF):

- Melange, a tool-supported meta-language to modularly define executable modeling languages with execution functions and data, and to extend (EMF-based) existing modeling languages. - MoCCML, a tool-supported meta-language dedicated to the specification of a Model of Concurrency and Communication (MoCC) and its mapping to a specific abstract syntax and associated execution functions of a modeling language. - GEL, a tool-supported meta-language dedicated to the specification of the protocol between the execution functions and the MoCC to support the feedback of the data as well as the callback of other expected execution functions. - BCOoL, a tool-supported meta-language dedicated to the specification of language coordination patterns to automatically coordinates the execution of, possibly heterogeneous, models. - Sirius Animator, an extension to the model editor designer Sirius to create graphical animators for executable modeling languages.

FUNCTIONAL DESCRIPTION: The GEMOC Studio is an eclipse package that contains components supporting the GEMOC methodology for building and composing executable Domain-Specific Modeling Languages (DSMLs). It includes the two workbenches: The GEMOC Language Workbench: intended to be used by language designers (aka domain experts), it allows to build and compose new executable DSMLs. The GEMOC Modeling Workbench: intended to be used by domain designersto create, execute and coordinate models conforming to executable DSMLs. The different concerns of a DSML, as defined with the tools of the language workbench, are automatically deployed into the modeling workbench. They parametrize a generic execution framework that provide various generic services such as graphical animation, debugging tools, trace and event managers, timeline, etc.

- Participants: Didier Vojtisek, Dorian Leroy, Erwan Bousse, Fabien Coulon and Julien Deantoni
- Partners: IRIT ENSTA I3S OBEO Thales TRT
- Contact: Benoît Combemale
- URL: http://gemoc.org/studio.html

### 6.4. Kevoree

### Kevoree Core

KEYWORDS: M2M - Dynamic components - Iot - Heterogeneity - Smart home - Cloud - Software architecture - Dynamic deployment

SCIENTIFIC DESCRIPTION: Kevoree is an open-source models@runtime platform (http://www.kevoree.org) to properly support the dynamic adaptation of distributed systems. Models@runtime basically pushes the idea of reflection [132] one step further by considering the reflection layer as a real model that can be uncoupled from the running architecture (e.g. for reasoning, validation, and simulation purposes) and later automatically resynchronized with its running instance.

Kevoree has been influenced by previous work that we carried out in the DiVA project [132] and the Entimid project [135]. With Kevoree we push our vision of models@runtime [131] farther. In particular, Kevoree provides a proper support for distributed models@runtime. To this aim we introduced the Node concept to model the infrastructure topology and the Group concept to model semantics of inter node communication during synchronization of the reflection model among nodes. Kevoree includes a Channel concept to allow for multiple communication semantics between remoteComponents deployed on heterogeneous nodes. All Kevoree concepts (Component, Channel, Node, Group) obey the object type design pattern to separate deployment artifacts from running artifacts. Kevoree supports multiple kinds of very different execution node technology (e.g. Java, Android, MiniCloud, FreeBSD, Arduino, ...).

Kevoree is distributed under the terms of the LGPL open source license.

Main competitors:

- the Fractal/Frascati eco-system (http://frascati.ow2.org ).
- SpringSource Dynamic Module (http://spring.io/)

GCM-Proactive (http://proactive.inria.fr/)

```
OSGi (http://www.osgi.org)
```

Chef

Vagran (http://vagrantup.com/)

Main innovative features:

distributed models@runtime platform (with a distributed reflection model and an extensible models@runtime dissemination set of strategies).

Support for heterogeneous node type (from Cyber Physical System with few resources until cloud computing infrastructure).

Fully automated provisioning model to correctly deploy software modules and their dependencies.

Communication and concurrency access between software modules expressed at the model level (not in the module implementation).

Impact:

Several tutorials and courses have been performed this year at EJCP for French PhD student, at ECNU summer school for 82 chineese PhD students. See also the web page http://www.kevoree.org.

In 2015, we mainly created a new implementation in C# and we created an implementation for system containers for driving resources using Kevoree. We also use Kevoree in the context of Mohammed's PhD to create testing infrastructure on-demand.

FUNCTIONAL DESCRIPTION: Kevoree is an open-source models@runtime platform to properly support the dynamic adaptation of distributed systems. Models@runtime basically pushes the idea of reflection one step further by considering the reflection layer as a real model that can be uncoupled from the running architecture (e.g. for reasoning, validation, and simulation purposes) and later automatically resynchronized with its running instance.

- Participants: Aymeric Hervieu, Benoit Baudry, Francisco-Javier Acosta Padilla, Inti Gonzalez Herrera, Ivan Paez Anaya, Jacky Bourgeois, Jean Emile Dartois, Johann Bourcier, Manuel Leduc, Maxime Tricoire, Mohamed Boussaa, Noël Plouzeau and Olivier Barais
- Contact: Olivier Barais
- URL: http://kevoree.org/

### 6.5. Melange

KEYWORDS: Modeling language - Meta-modelisation - Language workbench - Dedicated langage - Modeldriven software engineering - DSL - MDE - Meta model - Model-driven engineering - Meta-modeling

SCIENTIFIC DESCRIPTION: Melange is a follow-up of the executable metamodeling language Kermeta, which provides a tool-supported dedicated meta-language to safely assemble language modules, customize them and produce new DSMLs. Melange provides specific constructs to assemble together various abstract syntax and operational semantics artifacts into a DSML. DSMLs can then be used as first class entities to be reused, extended, restricted or adapted into other DSMLs. Melange relies on a particular model-oriented type system that provides model polymorphism and language substitutability, i.e. the possibility to manipulate a model through different interfaces and to define generic transformations that can be invoked on models written using different DSLs. Newly produced DSMLs are correct by construction, ready for production (i.e., the result can be deployed and used as-is), and reusable in a new assembly.

Melange is tightly integrated with the Eclipse Modeling Framework ecosystem and relies on the meta-language Ecore for the definition of the abstract syntax of DSLs. Executable meta-modeling is supported by weaving operational semantics defined with Xtend. Designers can thus easily design an interpreter for their DSL in a non-intrusive way. Melange is bundled as a set of Eclipse plug-ins.

FUNCTIONAL DESCRIPTION: Melange is a language workbench which helps language engineers to mashup their various language concerns as language design choices, to manage their variability, and support their reuse. It provides a modular and reusable approach for customizing, assembling and integrating DSMLs specifications and implementations.

- Participants: Arnaud Blouin, Benoît Combemale, David Mendez Acuna, Didier Vojtisek, Dorian Leroy, Erwan Bousse, Fabien Coulon, Jean-Marc Jézéquel, Olivier Barais and Thomas Degueule
- Contact: Benoît Combemale
- URL: http://melange-lang.org

### 6.6. Opencompare

KEYWORD: Software Product Line

FUNCTIONAL DESCRIPTION: Product comparison matrices (PCMs) are tabular data: supported and unsupported features are documented for both describing the product itself and for discriminating one product compared to another. PCMs abound – we are all using PCMs – and constitute a rich source of knowledge for easily comparing and choosing product. Yet the current practice is suboptimal both for humans and computers, mainly due to unclear semantics, heterogeneous forms of data, and lack of dedicated support.

OpenCompare.org is an ambitious project for the collaborative edition, the sharing, the standardisation, and the open exploitation of PCMs. The goal of OpenCompare.org is to provide an integrated set of tools (e.g., APIs, visualizations, configurators, editors) for democratizing their creation, import, maintenance, and exploitation.

- Participants: Guillaume Bécan, Mathieu Acher and Sana Ben Nasr
- Contact: Mathieu Acher
- URL: http://opencompare.org

## 7. New Results

### 7.1. Results on Variability modeling and management

### 7.1.1. Variability and testing.

Many approaches for testing configurable software systems start from the same assumption: it is impossible to test all configurations. This motivated the definition of variability-aware abstractions and sampling techniques to cope with large configuration spaces. Yet, there is no theoretical barrier that prevents the exhaustive testing of all configurations by simply enumerating them, if the effort required to do so remains acceptable. Not only this: we believe there is lots to be learned by systematically and exhaustively testing a configurable system. We report on the first ever endeavor to test all possible configurations of an industry-strength, open source configurable software system, JHipster, a popular code generator for web applications. We built a testing scaffold for the 26,000+ configurations of JHipster using a cluster of 80 machines during 4 nights for a total of 4,376 hours (182 days) CPU time. We find that 35.70% configurations fail and we identify the feature interactions that cause the errors. We show that sampling testing strategies (like dissimilarity and 2-wise) (1) are more effective to find faults than the 12 default configurations used in the JHipster continuous integration; (2) can be too costly and exceed the available testing budget. We cross this quantitative analysis with the qualitative assessment of JHipster's lead developers. Additional resources: preliminary effort on JHipster [32], https://arxiv.org/abs/1710.07980https://github.com/axel-halin/Thesis-JHipster/

### 7.1.2. Variability and teaching.

Software Product Line (SPL) engineering has emerged to provide the means to efficiently model, produce, and maintain multiple similar software variants, exploiting their common properties, and managing their variabilities (differences). With over two decades of existence, the community of SPL researchers and practitioners is thriving as can be attested by the extensive research output and the numerous successful industrial projects. Education has a key role to support the next generation of practitioners to build highly complex, variability-intensive systems. Yet, it is unclear how the concepts of variability and SPLs are taught, what are the possible missing gaps and difficulties faced, what are the benefits, or what is the material available. Also, it remains unclear whether scholars teach what is actually needed by industry. We report on three initiatives we have conducted with scholars, educators, industry practitioners, and students to further understand the connection between SPLs and education, i.e., an online survey on teaching SPLs we performed with 35 scholars, another survey on learning SPLs we conducted with 25 students, as well as two workshops held at the International Software Product Line Conference in 2014 and 2015 with both researchers and industry practitioners participating. We build upon the two surveys and the workshops to derive recommendations for educators to continue improving the state of practice of teaching SPLs, aimed at both individual educators as well as the wider community. Finally, we are developing and maintaining a repository for teaching SPLs and variability. Additional resources: https://teaching.variability.io

### 7.1.3. Variability and constraint solving.

Array constraints are essential for handling data structures in automated reasoning and software verification. Unfortunately, the use of a typical finite domain (FD) solver based on local consistency-based filtering has strong limitations when constraints on indexes are combined with constraints on array elements and size. This work proposes an efficient and complete FD-solving technique for extended constraints over (possibly unbounded) arrays. We describe a simple but particularly powerful transformation for building an equisatisfiable formula that can be efficiently solved using standard FD reasoning over arrays, even in the unbounded case. Experiments show that the proposed solver significantly outperforms FD solvers, and successfully competes with the best SMT-solvers [38]. This work is not directly related to variability and SPL. But it contributes to DiverSE's attempts to connect artificial intelligence techniques to software variability engineering, in which constraint solving or machine learning are typically applied.

## 7.1.4. Variability and machine learning (performance specialization of variability-intensive systems).

We propose the use of a machine learning approach to infer variability constraints from an oracle that is able to assess whether a given configuration is correct. We propose an automated procedure to randomly generate configurations, classify them according to the oracle, and synthesize cross-tree constraints. Specifically, based on an oracle (e.g. a runtime test) that tells us whether a given configuration meets the requirements (e.g. speed or memory footprint), we leverage machine learning to retrofit the acquired knowledge into a variability model of the system that can be used to automatically specialize the configurable system. We validate our approach on a set of well-known configurable software systems (Apache server, x264, etc.) Our results show that, for many different kinds of objectives and performance qualities, the approach has interesting accuracy, precision and recall after a learning stage based on a relative small number of random samples [43]. Additional resources: https://learningconstraints.github.io and VaryVary ANR project

### 7.1.5. Variability and machine learning (learning contextual variability models).

Modeling how contextual factors relate to a software system's configuration space is usually a manual, errorprone task that depends highly on expert knowledge. Machine-learning techniques can automatically predict the acceptable software configurations for a given context. Such an approach executes and observes a sample of software configurations within a sample of contexts. It then learns what factors of each context will likely discard or activate some of the software's features. This lets developers and product managers automatically extract the rules that specialize highly configurable systems for specific contexts [27] Additional resources: https://learningconstraints.github.io and VaryVary ANR project

We are currently exploring the use of machine learning for variability-intensive systems in the context of VaryVary ANR project (see also VaryLaTeX [28]).

### 7.2. Results on Software Language Engineering

### 7.2.1. On Language Interfaces

Complex systems are developed by teams of experts from multiple domains , who can be liberated from becoming programming experts through domain-specific languages (DSLs). The implementation of the different concerns of DSLs (including syntaxes and semantics) is now well-established and supported by various languages workbenches. However, the various services associated to a DSL (e.g., editors, model checker, debugger or composition operators) are still directly based on its implementation. Moreover, while most of the services crosscut the different DSL concerns, they only require specific information on each. Consequently, this prevents the reuse of services among related DSLs, and increases the complexity of service implementation. Leveraging the time-honored concept of interface in software engineering, we discuss in [40] the benefits of language interfaces in the context of software language engineering. In particular, we elaborate on particular usages that address current challenges in language development.

### 7.2.2. Revisiting Visitors for Modular Extension of Executable DSMLs

Executable Domain-Specific Modeling Languages (xDSMLs) are typically defined by metamodels that specify their abstract syntax, and model interpreters or compilers that define their execution semantics. To face the proliferation of xDSMLs in many domains, it is important to provide language engineering facilities for opportunistic reuse, extension, and customization of existing xDSMLs to ease the definition of new ones. Current approaches to language reuse either require to anticipate reuse, make use of advanced features that are not widely available in programming languages, or are not directly applicable to metamodel-based xDSMLs. In [35], we propose a new language implementation pattern, named REVISITOR, that enables independent extensibility of the syntax and semantics of metamodel-based xDSMLs with incremental compilation and without anticipation. We seamlessly implement our approach alongside the compilation chain of the Eclipse Modeling Framework, thereby demonstrating that it is directly and broadly applicable in various modeling environments. We show how it can be employed to incrementally extend both the syntax and semantics of the fUML language without requiring anticipation or re-compilation of existing code, and with acceptable performance penalty compared to classical handmade visitors.

## 7.2.3. Advanced and efficient execution trace management for executable domain-specific modeling languages

Executable Domain-Specific Modeling Languages (xDSMLs) enable the application of early dynamic verification and validation (V&V) techniques for behavioral models. At the core of such techniques, execution traces are used to represent the evolution of models during their execution. In order to construct execution traces for any xDSML, generic trace metamodels can be used. Yet, regarding trace manipulations, generic trace metamodels lack efficiency in time because of their sequential structure, efficiency in memory because they capture superfluous data, and usability because of their conceptual gap with the considered xDSML. We contributed in [22] a novel generative approach that defines a multidimensional and domain-specific trace metamodel enabling the construction and manipulation of execution traces for models conforming to a given xDSML. Efficiency in time is improved by providing a variety of navigation paths within traces, while usability and memory are improved by narrowing the scope of trace metamodels to fit the considered xDSML. We evaluated our approach by generating a trace metamodel for fUML and using it for semantic differencing, which is an important V&V technique in the realm of model evolution. Results show a significant performance improvement and simplification of the semantic differencing rules as compared to the usage of a generic trace metamodel.

### 7.2.4. Omniscient Debugging for Executable DSLs

Omniscient debugging is a promising technique that relies on execution traces to enable free traversal of the states reached by a model (or program) during an execution. While a few General-Purpose Languages (GPLs) already have support for omniscient debugging, developing such a complex tool for any executable Domain Specific Language (DSL) remains a challenging and error prone task. A generic solution must: support a wide range of executable DSLs independently of the metaprogramming approaches used for implementing their semantics; be efficient for good responsiveness. Our contribution in [21] relies on a generic omniscient debugger supported by efficient generic trace management facilities. To support a wide range of executable DSLs, the debugger provides a common set of debugging facilities, and is based on a pattern to define runtime services independently of metaprogramming approaches. Results show that our debugger can be used with various executable DSLs implemented with different metaprogramming approaches. As compared to a solution that copies the model at each step, it is on average six times more efficient in memory, and at least 2.2 faster when exploring past execution states, while only slowing down the execution 1.6 times on average.

### 7.2.5. Reverse Engineering Language Product Lines from Existing DSL Variants

The use of domain-specific languages (DSLs) has become a successful technique in the development of complex systems. Nevertheless, the construction of this type of languages is time-consuming and requires highly-specialized knowledge and skills. An emerging practice to facilitate this task is to enable reuse through the definition of language modules which can be later put together to build up new DSLs. In [26], we

propose a reverse-engineering technique to ease-off such a development scenario. Our approach receives a set of DSL variants which are used to automatically recover a language modular design and to synthesize the corresponding variability models. The validation is performed in a project involving industrial partners that required three different variants of a DSL for finite state machines. This validation shows that our approach is able to correctly identify commonalities and variability.

### 7.2.6. Software Language Engineering for Virtual Reality Software Development

Due to the nature of Virtual Reality (VR) research, conducting experiments in order to validate the researcher's hypotheses is a must. However, the development of such experiments is a tedious and time-consuming task. In [48], we propose to make this task easier, more intuitive and faster with a method able to describe and generate the most tedious components of VR experiments. The main objective is to let experiment designers focus on their core tasks: designing , conducting, and reporting experiments. To that end, we applied well-established SLE concepts promoted in DIVERSE to the VR domain to ease the development of VR experiments. More precisely, we propose the use of DSLs to ease the description and generation of VR experiments. An analysis of published VR experiments is used to identify the main properties that characterize VR experiments. This allowed us to design AGENT (Automatic Generation of ExperimeNtal proTocol runtime), a DSL for specifying and generating experimental protocol runtimes. We demonstrated the feasibility of our approach by using AGENT on two experiments published in the VRST'16 proceedings.

### 7.2.7. Create and Play your Pac-Man Game with the GEMOC Studio

Executable Domain-Specific Languages (DSLs) are used for defining the behaviors of systems. In particular, the operational semantics of such DSLs may define how conforming models react to stimuli from their environment. This commonly requires adapting the semantics to define both the possible domain-level stimuli, and their handling during the execution. However, manually adapting the semantics for such cross-cutting concern is a complex and error-prone task. In , we demonstrate a tool addressing this problem by allowing the augmentation of operational semantics for handling stimuli, and by automatically generating a complete behavioral language interface from this augmentation. At runtime, this interface can receive stimuli sent to models, and can safely handle them by automatically interrupting the execution flow. This tool is an extension to the GEMOC Studio, a language and modeling workbench for executable DSLs We demonstrate how it can be used to implement a Pac-Man DSL enabling to create and play Pac-Man games.

### 7.3. Results on Heterogeneous and dynamic software architectures

We have selected three main contributions for DIVERSE's research axis #4: one is in the field of runtime management, while the two others one are in the field of Privacy and Security.

## 7.3.1. Verifying the configuration of Virtualized Network Functions in Software Defined Networks

In Kevoree, one of the goal is to work on the shipping pases in which we aim at making deployment, and the reconfiguration simple and accessible to the whole team. This year we work to include the capacity to manage network configuration when reconfiguring application stack. In this context, the deployment of modular virtual network functions (VNFs) in software defined infrastructures (SDI) enables cloud and network providers to deploy integrated network services across different resource domains. It leads to a large interleaving between network configuration through software defined network controllers and VNF deployment within this network. Most of the configuration management tools and network orchestrator used to deploy VNF lack of an abstraction to express Assume-Guarantee contracts between the VNF and the SDN configuration. Consequently, VNF deployment can be inconsistent with network configurations.

Contribution. To tackle this challenge, in this work [41], we develop an approach to check the consistency between the VNF description described from a set of structural models and flow-chart models and a proposed deployment on a real SDN infrastructure with its own configuration manager. We illustrate our approach on virtualized Evolved Packet Core function.

Originality. The originality of this work is to propose a model to capture VNF.

Impact. Beyond the scientific originality of this work, the main impacts of this novel approach to check SDN configuration has been to (i) reinforce DIVERSE's visibility in the academic and industrial communities on software components and (ii) to create several research tracks that are currently explored in different projects of the team (B-com PhD thesis and Nokia common labs). This work is being integrated within the Kevoree platform.

### 7.3.2. Identity Negotiation at Runtime

Authentication delegation is a major function of the modern web. Identity Providers (IdP) acquired a central role by providing this function to other web services. By knowing which web services or web applications access its service, an IdP can violate the end-user privacy by discovering information that the user did not want to share with its IdP. For instance, WebRTC introduces a new field of usage as authentication delegation happens during the call session establishment, between two users. As a result, an IdP can easily discover that Bob has a meeting with Alice. A second issue that increases the privacy violation is the lack of choice for the end-user to select its own IdP. Indeed, on many web-applications, the end-user can only select between a subset of IdPs, in most cases Facebook or Google.

- Contribution. This year, we analyze this phenomena [23], in particular why the end-user cannot easily select its preferred IdP, though there exists standards in this field such as OpenID Connect and OAuth 2? To lead this analysis, we conduct three investigations. The first one is a field survey on OAuth 2 and OpenID Connect scope usage by web sites to understand if scopes requested by web-sites could allow for user defined IdPs. The second one tries to understand whether the problem comes from the OAuth 2 protocol or its implementations by IdP. The last one tries to understand if trust relations between websites and IdP could prevent the end user to select its own IdP. Finally, we sketch possible architecture for web browser based identity management, and report on the implementation of a prototype. We also describe our implementation of the WebRTC identity architecture [24]. We adapt OpenID Connect servers to support WebRTC peer to peer authentication and detail the issues and solutions found in the process.
- Originality. We observe that although WebRTC allows for the exchange of identity assertion between peers, users lack feedback and control over the other party authentication. To allow identity negotiation during a WebRTC communication setup, we propose an extension to the Session Description Protocol. Our implementation demonstrates current limitations with respect to the current WebRTC specification.

Impact. This work is done with Orange.

### 7.3.3. Raising Time Awareness in Model-Driven Engineering

The conviction that big data analytics is a key for the success of modern businesses is growing deeper, and the mobilisation of companies into adopting it becomes increasingly important. Big data integration projects enable companies to capture their relevant data, to efficiently store it, turn it into domain knowledge, and finally monetize it. In this context, historical data, also called temporal data, is becoming increasingly available and delivers means to analyse the history of applications, discover temporal patterns, and predict future trends. Despite the fact that most data that today's applications are dealing with is inherently temporal current approaches, methodologies, and environments for developing these applications don't provide sufficient support for handling time. We envision that Model-Driven Engineering (MDE) would be an appropriate ecosystem for a seamless and orthogonal integration of time into domain modeling and processing.

- Contribution. This year, we investigate the state-of-the-art in MDE techniques and tools in order to identify the missing bricks for raising time-awareness in MDE and outline research directions in this emerging domain [30].
- Originality. We propose an extended context representation for self-adaptive software that integrates the history of planned actions as well as their expected effects over time into the context representations. We demonstrate on a cloud elasticity manager case study that such *temporal action-aware context*

leads to improved reasoners while still be highly scalable. This work is original with respect to the state of the art since it provides a way to represent and take into account the impact of reconfiguration actions on a system.

Impact. This work is done through a collaboration with the SnT in Luxembourg and a startup called DataThings, working on domain model representation for various industrial domains.

### 7.3.4. Collaborations

This year, we had a close and fruitful collaboration with the industrial partners that are involved in the HEADS and Occiware projects, in particular an active interaction with the Tellu company in Norway in the Heads context. Tellu relies on Kevoree and KevoreeJS to build their health management systems. They will be also an active member the new Stamp project led by DIVERSE. We can cite also an active collaboration with Orange Labs through Kevin Corre's joint PhD thesis. Another joint industrial (CIFRE) PhD started in September 2016, and we are also partner in a new starting FUI project. Finally, DIVERSE collaborates with the B-COM IRT (https://b-com.com/en), as one permanent member has a researcher position of one day per week at B-COM and a new joint PhD started in September.

At the academic level we collaborate actively with the Spiral team at Inria Lille (several joint projects), the Tacoma team (with two co-advised PhD students), the Myriad team (1 co-advised PhD student) and we have started two collaborations with the ASAP team.

### 7.4. Results on Diverse Implementations for Resilience

Diversity is acknowledged as a crucial element for resilience, sustainability and increased wealth in many domains such as sociology, economy and ecology. Yet, despite the large body of theoretical and experimental science that emphasizes the need to conserve high levels of diversity in complex systems, the limited amount of diversity in software-intensive systems is a major issue. This is particularly critical as these systems integrate multiple concerns, are connected to the physical world, run eternally and are open to other services and to users. Here we present our latest observational and technical results about (i) new approaches to increase diversity in software systems, and (ii) software testing to assess the validity of software.

### 7.4.1. Software diversification

Our work on software diversification explores various ways of adding randomness in program executions: state perturbations that preserve functional correctness [25]; randomizing of web APIs to mitigate browser fingeprinting [33].

Can the execution of software be perturbed without breaking the correctness of the output? In this work [25], we devise a protocol to answer this question from a novel perspective. In an experimental study, we observe that many perturbations do not break the correctness in ten subject programs. We call this phenomenon "correctness attraction". The uniqueness of this protocol is that it considers a systematic exploration of the perturbation space as well as perfect oracles to determine the correctness of the output. To this extent, our findings on the stability of software under execution perturbations have a level of validity that has never been reported before in the scarce related work. A qualitative manual analysis enables us to set up the first taxonomy ever of the reasons behind correctness attraction.

The rich programming interfaces (APIs) provided by web browsers can be diverted to collect a browser fingerprint. A small number of queries on these interfaces are sufficient to build a fingerprint that is statistically unique and very stable over time. Consequently, the fingerprint can be used to track users. Our work [33] aims at mitigating the risk of browser fingerprinting for users privacy by 'breaking' the stability of a fingerprint over time. We add randomness in the computation of selected browser functions, in order to have them deliver slightly different answers for each browsing session. Randomization is possible thanks to the following properties of browsers implementations: (i) some functions have a nondeterministic specification, but a deterministic implementation ; (ii) multimedia functions can be slightly altered without deteriorating user's perception. We present FPRandom, a modified version of Firefox that adds randomness to mitigate the most recent fingerprinting algorithms, namely canvas fingerprinting, AudioContext fingerprinting and the

unmasking of browsers through the order of JavaScript properties. We evaluate the effectiveness of FPRandom by testing it against known fingerprinting tests. We also conduct a user study and evaluate the performance overhead of randomization to determine the impact on the user experience.

The other aspect in the area of software diversity is about the statistical analysis of browser fingerprinting on a large industrial dataset [17], [31].

### 7.4.2. Software testing

Generative software development has paved the way for the creation of multiple code generators and compilers that serve as a basis for automatically generating code to a broad range of software and hardware platforms. With full automatic code generation, the user is able to easily and rapidly synthetize software artifacts for various software platforms. In addition, modern generators (i.e., C compilers) become highly configurable, offering numerous configuration options that the user can use to easily customize the generated code for the target hardware platform. In this context, it is crucial to verify the correct behaviour of code generators. Numerous approaches have been proposed to verify the functional outcome of generated code but few of them evaluate the non-functional properties of automatically generated code, namely the performance and resource usage properties. The thesis of Mohamed Boussaa [16] has addressed this limitation.

## 8. Bilateral Contracts and Grants with Industry

### 8.1. Bilateral Contracts with Industry

### 8.1.1. GLOSE

- Partners: Inria/CNRS/Safran
- Dates: 2017-2021
- Abstract: The GLOSE project develops new techniques for heterogeneous modeling and simulation in the context of systems engineering. It aims to provide formal and operational tools and methods to formalize the behavioral semantics of the various modeling languages used at system-level. These semantics will be used to extract behavioral language interfaces supporting the definition of coordination patterns. These patterns, in turn, can systematically be used to drive the coordination of any model conforming to these languages. The project is structured according to the following tasks: concurrent xDSML engineering, coordination of discrete models, and coordination of discrete/continuous models. The project is funded in the context of the network DESIR, and supported by the GEMOC initiative.

### 8.1.2. One Shot Software (OSS)

- Partners: Inria/Orange
- Dates: 2017-2019
- Abstract: The OSS project investigates an extreme version of moving target defense where a slightly different version of the application is deployed each time it is used (e.g., for crypto functions or payment services). We investigate the analysis, synthesis and transformation techniques to support diversification at 5 points of a software construction pipeline, which, once combined yield up to billions of variants. We also evaluate the support of diversification as a first class property in DevOps.

## 9. Partnerships and Cooperations

### 9.1. National Initiatives

**9.1.1. ANR** 9.1.1.1. SOPRANO

- Coordinator: CEA
- CEA, University of Paris-Sud, Inria Rennes, OcamlPro, Adacore
- Dates: 2014-2017
- Abstract: Today most major verification approaches rely on automatic external solvers. However these solvers do not fill the current and future needs for verification: lack of satisfying model generation, lack of reasoning on difficult theories (e.g. floating-point arithmetic), lack of extensibility for specific or new needs. The SOPRANO project aims at solving these problems and prepare the next generation of verification-oriented solvers by gathering experts from academia and industry. We will design a new framework for the cooperation of solvers, focused on model generation and borrowing principles from SMT (current standard) and CP (well-known in optimisation). These ideas will be implemented in an open-source platform, with regular evaluations from the industrial partners.

### 9.1.1.2. VaryVary ANR JCJC

- Coordinator: Mathieu Acher
- DiverSE, Inria/IRISA Rennes
- Dates: 2017-2021
- Abstract: Most modern software systems (operating systems like Linux, Web browsers like Firefox • or Chrome, video encoders like x264 or ffmpeg, servers, mobile applications, etc.) are subject to variation or come in many variants. Hundreds of configuration options, features, or plugins can be combined, each potentially with distinct functionality and effects on execution time, memory footprint, etc. Among configurations, some of them are chosen and do not compile, crash at runtime, do not pass a test suite, or do not reach a certain performance quality (e.g., energy consumption, security). In this JCJC ANR project, we follow a thought-provocative and unexplored direction: We consider that the variability boundary of a software system can be specialized and should vary when needs be. The goal of this project is to provide theories, methods and techniques to make vary variability. Specifically, we consider machine learning and software engineering techniques for narrowing the space of possible configurations to a good approximation of those satisfying the needs of users. Based on an oracle (e.g., a runtime test) that tells us whether a given configuration meets the requirements (e.g. speed or memory footprint), we leverage machine learning to retrofit the acquired constraints into a variability that can be used to automatically specialize the configurable system. Based on a relative small number of configuration samples, we expect to reach high accuracy for many different kinds of oracles and subject systems. Our preliminary experiments suggest that varying variability can be practically useful and effective. However, much more work is needed to investigate sampling, testing, and learning techniques within a variety of cases and application scenarios. We plan to further collect large experimental data and apply our techniques on popular, open-source, configurable software (like Linux, Firefox, ffmpeg, VLC, Apache or JHipster) and generators for media content (like videos, models for 3D printing, or technical papers written in LaTeX).

### 9.1.1.3. CLARITY

- Coordinator: Obéo
- Other partners: AIRBUS, Airbus Defence and Space, All4tec, ALTRAN Technologies, AREVA, Artal, C.E.S.A.M.E.S., Eclipse Foundation Europe, Inria Sophia Antipolis Méditerranée, PRFC, Scilab Enterprises, Thales Global Services, Thales Alenia Space, Thales Research & Technology, Thales Systèmes Aéroportés, Université de Rennes 1.
- Dates: 2014-2017
- Abstract: The CLARITY project aims to establish an international dimension ecosystem around Melody/Capella modeling workbench for systems engineering (MBSE) and engineering architectures (system, software, hardware).

### 9.1.1.4. Occiware

- Coordinator: Open Wide
- Open Wide, ActiveEon SA, CSRT Cloud Systèmes Réseaux et Télécoms, Institut Mines-Télécom/Télécom SudParis, Inria, Linagora, Obeo, OW2 Consortium, Pôle Numérique, Université Joseph Fourier,
- Dates: 2014-2017
- Abstract: The Occiware project aims to establish a formal and equipped framework for the management of all cloud resource based on the OCCI standard.

### 9.1.2. DGA

### 9.1.2.1. FPML (CYBERDEFENSE)

- Coordinator: DGA
- Partners: DGA MI, Inria
- Dates: 2014-2017
- Abstract: in the context of this project, DGA-MI and the Inria team DiverSE explore the existing approaches to ease the development of formal specifications of domain-Specific Languages (DSLs) dedicated to paquet filtering, while guaranteeing expressiveness, precision and safety. In the long term, this work is part of the trend to provide to DGA-MI and its partners a tooling to design and develop formal DSLs which ease the use while ensuring a high level of reasoning.

### 9.1.3. Cominlabs

9.1.3.1. PROFILE

- Coordinator: Université de Rennes 1
- Partners: Inria, Université de Rennes 2
- Dates: 2016-2019
- Abstract: The PROFILE project brings together experts from law, computer science and sociology to address the challenges raised by online profiling, following a multidisciplinary approach. More precisely, the project will pursue two complementary and mutually informed lines of research: (i) Investigate, design, and introduce a new right of opposition into the legal framework of data protection to better regulate profiling and to modify the behavior of commercial companies towards being more respectful of the privacy of their users; (ii)S Provide users with the technical means they need to detect stealthy profiling techniques as well as to control the extent of the digital traces they routinely produce. As a case study, we focus on browser fingerprinting, a new profiling technique for targeted advertisement. The project will develop a generic framework to reason on the data collected by profiling algorithms, to uncover their inner working, and make them more accountable to users. PROFILE will also propose an innovative protection to mitigate browser fingerprinting, based on the collaborative reconfiguration of browsers.

### 9.2. European Initiatives

### 9.2.1. FP7 & H2020 Projects

### 9.2.1.1. FP7 STREP HEADS

- Coordinator: SINTEF
- Other partners: Inria, Software AG, ATC, Tellu, eZmonitoring
- Dates: 2013-2017

• Abstract: The idea of the HEADS project is to leverage model-driven software engineering and generative programming techniques to provide a new integrated software engineering approach which allow advanced exploitation the full range of diversity and specificity of the future computing continuum. The goal is to empower the software and services industry to better take advantage of the opportunities of the future computing continuum and to effectively provide new innovative services that are seamlessly integrated to the physical world making them more pervasive, more robust, more reactive and closer (physically, socially, emotionally, etc.) to their users. We denote such services or applications within the Future Internet whose logic and value emerges from a set of communicating software components distributed on a heterogeneous computing continuum from clouds to mobile devices, sensors and/or smart-objects.

### 9.2.1.2. H2020 ICT-10-2016 STAMP

- Coordinator: Inria Rennes
- Other partners: ATOS, ActiveEon, OW2, TellU, Engineering, XWiki, TU Delft, SINTEF
- Dates: 2016-2019
- Abstract: Leveraging advanced research in automatic test generation, STAMP aims at pushing automation in DevOps one step further through innovative methods of test amplification. It will reuse existing assets (test cases, API descriptions, dependency models), in order to generate more test cases and test configurations each time the application is updated. Acting at all steps of development cycle, STAMP techniques aim at reducing the number and cost of regression bugs at unit level, configuration level and production stage.

STAMP will raise confidence and foster adoption of DevOps by the European IT industry. The project gathers 3 academic partners with strong software testing expertise, 5 software companies (in: e-Health, Content Management, Smart Cities and Public Administration), and an open source consortium. This industry-near research addresses concrete, business-oriented objectives. All solutions are open source and developed as microservices to facilitate exploitation, with a target at TRL 6.

### 9.2.2. Collaborations with Major European Organizations

- SINTEF, ICT (Norway): Model-driven systems development for the construction of distributed, heterogeneous applications. We collaborate since 2008 and are currently in two FP7 projects together.
- Université du Luxembourg, (Luxembourg): Models runtime for dynamic adaptation and multiobjective elasticity in cloud management; model-driven development.
- KTH, the Royal Institute of Technology (Sweden): continuous software testing, perturbation and diversification.

### 9.3. International Initiatives

### 9.3.1. Inria Associate Teams Not Involved in an Inria International Labs

9.3.1.1. ALE

- Title: Agile Language Engineering
- International Partner (Institution Laboratory Researcher):
  - CWI (Netherlands)
- Start year: 2017
- See also: http://gemoc.org/ale/

Software engineering faces new challenges with the advent of modern software-intensive systems such as complex critical embedded systems, cyber-physical systems and the Internet of things. Application domains range from robotics, transportation systems, defense to home automation, smart cities, and energy management, among others. Software is more and more pervasive, integrated into large and distributed systems, and dynamically adaptable in response to a complex and open environment. As a major consequence, the engineering of such systems involves multiple stakeholders, each with some form of domain-specific knowledge, and with an increasingly use of software as an integration layer. Hence more and more organizations are adopting Domain Specific Languages (DSLs) to allow domain experts to express solutions directly in terms of relevant domain concepts. This new trend raises new challenges about designing DSLs, evolving a set of DSLs and coordinating the use of multiple DSLs for both DSL designers and DSL users. ALE will contribute to the field of Software Language Engineering, aiming to provide more agility to both language designers and language users. The main objective is twofold. First, we aim to help language designers to leverage previous DSL implementation efforts by reusing and combining existing language modules. Second, we aim to provide more flexibility to language users by ensuring interoperability between different DSLs and offering live feedback about how the model or program behaves while it is being edited (aka. live programming/modeling).

### 9.3.2. Inria International Partners

### 9.3.2.1. Informal International Partners

- Université de Montréal (Canada)
- McGill University (Canada)
- University of Alabama (USA)
- TU Wien (Austria)
- Michigan State University (MSU)
- Aachen University (Germany)
- KTH (Sweden)

### 9.3.3. Participation in Other International Programs

The GEMOC studio has been sustained through the creation of a Research Consortium at the Eclipse Foundation.

### 9.3.4. International initiative GEMOC

The GEMOC initiative (cf. http://www.gemoc.org) is an open and international initiative launched in 2013 that coordinate research partners worldwide to develop breakthrough software language engineering (SLE) approaches that support global software engineering through the use of multiple domain-specific languages. GEMOC members aim to provide effective SLE solutions to problems associated with the design and implementation of collaborative, interoperable and composable modeling languages.

The GEMOC initiative aims to provide a framework that facilitates collaborative work on the challenges of using of multiple domain-specific languages in software development projects. The framework consists of mechanisms for coordinating the work of members, and for disseminating research results and other related information on GEMOC activities. The framework also provides the required infrastructure for sharing artifacts produced by members, including publications, case studies, and tools.

The governance of the GEMOC initiative is ensured by the Advisory Board. The role of the Advisory Board is to coordinate the GEMOC work and to ensure proper dissemination of work products and information about GEMOC events (e.g., meetings, workshops).

Benoit Combemale is the co-founder and currently acts as principal coordinator of the GEMOC initiative. Benoit Combemale and Jean-Marc Jézéquel are part of the Advisory Board, and 9 DIVERSE members are part of the GEMOC initiative.

### 9.4. International Research Visitors

### 9.4.1. Visits of International Scientists

Yves Le Traon, Professor at the University of Luxembourg, visited the team in June and July 2017.

Tanja Mayerhofer, Junior Researcher at the TU Wien, visited the team in March 2017.

François Fouquet, Junior Researcher at the SnT (Lux), visited the team in November 2017.

### 9.4.1.1. Internships

Koko armando Nguepi kenfack, Master interships at the University of Namur, visited the team from September 2017 to January 2018.

### 9.4.2. Visits to International Teams

Manuel Leduc visited CWI for 3 weeks in September 2017

Benoit Combemale visited Professor Jorg Kienzle at McGill University (Canada) for 3 months in 2017; and made several short visits at CWI (The Netherlands).

### 9.4.2.1. Research Stays Abroad

Marcelino Rodriguez-Cancio visited Vanderbildt University from November 2016 to September 2017.

## **10. Dissemination**

### **10.1. Promoting Scientific Activities**

### 10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- Benoit Combemale has been general chair for SLE 2017, the major international conference in the area of software language engineering.
- Benoit Combemale has been main organizer of the Dagstuhl Seminar #17342 on "The Software Language Engineering Body of Knowledge" (SLEBoK).
- Mathieu Acher has been program committee co-chair of SPLC 2017, the major international conference in the area of variability and software product line engineering.
- 10.1.1.2. Member of the Organizing Committees

Arnaud Blouin:

• Publicity co-chair, EICS'17, 2017

Mathieu Acher:

- Publicity chair, ICSR'18
- Co-organizer of REVE'17 (international workshop on reverse engineering variability)

### 10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

Mathieu Acher:

- PC member ICSR 2017
- PC member VaMoS 2017
- PC member SAC 2017
- PC member SEAA 2017
- PC member GramSec'17
Olivier Barais:

- PC member SEAA 2017
- PC member SAC 2017
- PC member ICWE 2017

Benoit Baudry:

- PC member ICSE 2017
- PC member ASE 2017

Arnaud Blouin:

• PC member of the ACM Student Research Competition (SRC) at MODELS 2017

Benoit Combemale:

- PC member for MODELS'17
- PC member for ICMT'17
- PC member for the GEMOC'17 workshop at MODELS'17
- PC member for the MDEbug'17 workshop at MODELS'17
- PC member for the EXE'17 workshop at MODELS'17
- PC member for the MiSE'17 workshop at ICSE'17
- PC member for the MoMo'17 workshop at Modularity'17

Jean-Marc Jézéquel:

- PC member ICSE 2017
- PC member SPLC 2017 Vision track
- PC member SEAMS 2017

Johann Bourcier:

- PC member for the SEsCPS'17 workshop at ICSE'17
- 10.1.2.2. Reviewer

Arnaud Blouin was an external reviewer for ICSE 2017, EICS 2017, MODELS 2017, ICSA 2017, SEAMS 2017. Olivier Barais was a reviewer for ICSE 2017, Sosym 2017, JSS, JISA.

Johann Bourcier was a reviewer for ASE 2017, ICSE 2017 and Models 2017.

### 10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Benoit Baudry:

- SOSYM
- STVR

Benoit Combemale:

- Springer Journal on Software and System Modeling (SoSYM),
- Elsevier Journal on Computer Languages, Systems and Structures (COMLAN),
- Elsevier Journal on Science of Computer Programming (SCP), Advisory Board member of the Software Section.

Jean-Marc Jézéquel is Associate Editor in Chief of the Journal of Software and Systems Modeling: SoSyM, and member of the Editorial Board of:

- IEEE Computer
- Journal of Systems and Software: JSS
- Journal of Object Technology: JOT

### 10.1.3.2. Reviewer - Reviewing Activities

Arnaud Blouin: JSS, Ingénierie des Systèmes d'Information

Johann Bourcier was reviewer for the following magazine : IEEE Communication Magazine.

### 10.1.4. Invited Talks

Benoit Baudry:

- Software technology and DevOps. Scientific days of Orange
- Reconciling Diversity and Privacy at the Dagstuhl Seminar on Online Privacy and Web Transparency
- Software diversification as an obfuscation technique, at the International Workshop on Obfuscation: Science, Technology, and Theory

Benoit Combemale:

- From Model (driven) Engineering, to Language (driven) Engineering. Invited talk at CEA DAM, France.
- Towards Language-Oriented Modeling. Invited talk at UQAM (Montreal, Canada).
- Sound, yet Flexible, Modeling: A Language Engineering Point Of View. Keynote at FlexMDE'17.
- Modeling For Sustainability Or How to Make Smart CPS Smarter? Invited talk at CWI (Amsterdam, The Netherlands).
- Model Simulation, Graphical Animation, and Omniscient Debugging with Sirius Animator. Invited talk at SiriusCon'17.

### 10.1.5. Leadership within the Scientific Community

Benoit Baudry:

• Steering committee member for the ACM/IEEE MODELS conference

Benoit Combemale:

- Steering committee member for the ACM SLE conference
- Founding member and member of the advisory board of the GEMOC initiative.

Arnaud Blouin:

• Founding member and member of the GL-IHM (software engineering and human-computer interaction) working group (action spécifique GDR-GPL 2017)

### 10.1.6. Scientific Expertise

Arnaud Blouin: external ANR (French national research agency) reviewer 2017 Olivier Barais: International collaborations foreign minister. Expert Olivier Barais: AutoActive proposal board of expert (Norvegian project) Olivier Barais: AutoActive proposal board of expert (Norvegian project)

Johann Bourcier : scientific reviewer for CIR (Credit Impot Recherche).

### 10.1.7. Research Administration

Benoit Baudry is in the scientific advisory board of the SVV lab, University of Luxembourg.

Jean-Marc Jézéquel is the Director of IRISA, a 800+ people joint labs of CNRS, ENS Rennes, INSA Rennes, Inria, Telecom Bretagne, CentraleSupelec, the University of Rennes 1 and the University of South Brittany. He is also the head of Research of the French CyberSecurity Excellence Cluster, as well as the EIT Digital Rennes Satellite Node Director.

### 10.2. Teaching - Supervision - Juries

### 10.2.1. Teaching

The DIVERSE team bears the bulk of the teaching on Software Engineering at the University of Rennes 1 and at INSA Rennes, for the first year of the Master of Computer Science (Project Management, Object-Oriented Analysis and Design with UML, Design Patterns, Component Architectures and Frameworks, Validation & Verification, Human-Computer Interaction) and for the second year of the MSc in software engineering (Model driven Engineering, Aspect-Oriented Software Development, Software Product Lines, Component Based Software Development, Validation & Verification, *etc.*).

Each of Jean-Marc Jézéquel, Noël Plouzeau, Olivier Barais, Johann Bourcier, Arnaud Blouin, Mathieu Acher and Benoit Combemale teaches about 200h in these domains, with Benoit Baudry and teaching about 50h, for a grand total of about 1500 hours, including several courses at ENSTB, Supelec, and ENSAI Rennes engineering school.

Olivier Barais is deputy director of the electronics and computer science teaching department of the University of Rennes 1. Mathieu Acher is in charge of teaching duties management of this department. Noël Plouzeau is the head of the final year of the Master in Computer Science at the University of Rennes 1. Johann Bourcier is at the board of the ESIR engineering school.

The DIVERSE team also hosts several MSc and summer trainees every year.

### 10.2.2. Supervision

- PhD : Mohamed Boussaa, Automatic non-functional testing and tuning of configurable generators, 06/09/2017, B. Baudry, O. Barais
- PhD : Marcelino Rodriguez Cancio, *Automatic computation diversification*, 19/12/2017, B. Baudry and B. Combemale
- PhD: Pierre Laperdrix, Browser Fingerprinting: Exploring Device Diversity to Augment Authentication and Build Client-Side Countermeasures, 03/11/17, B. Baudry, G. Avoine
- PhD in progress: Alejandro Gomez Boix, *Distributed counter-measure against browser fingerprinting*, 2016, B. Baudry, D. Bromberg
- PhD in progress: Nicolas Harrand, Automatic diversity for code obfuscation, 2016, B. Baudry
- PhD in progress: Johan Pelay, *Langage pour une programmation incrémentale de réseau*, 2016, O. Barais, F. Guillemin
- PhD in progress: Quentin Plazar, *Bridging the gap between SAT and SNP*, 2015, M. Acher, A. Goetib
- PhD in progress: Paul Temple, Leraning variability models, 2015, M. Acher, J.-M. Jézéquel
- PhD in progress: Manuel Leduc, *Formal and Executable Specification of domain-specific language families*, 2016, O. Barais, B. Combemale
- PhD in progress: Youssou NDiaye, *Modelling and evaluating security in user interfaces*, 2016, N. Aillery, O. Barais, A. Blouin, A. Bouabdallah
- PhD in progress: Jean-Émile Dartois, *Performance resources modeling for container based applications*, 2016, O. Barais
- PhD in progress: Oscar Luis, Automatic test amplification, 2016, B. Baudry
- PhD in progress: Alexandre Rio, *Activity modeling for better renewable energy usage*, 2016, O. Barais, Y. Morel
- PhD: in progress, Kevin Corre, Modélisation de la confiance dans les services sociaux et conversationnels, 2014, O. Barais, G. Sunye
- PhD: in progress, Gwendal Le Moulec, *Towards the automatic synthesis of virtual reality applications*, 2015, B. Arnaldi, A. Blouin, V. Gouranton

- PhD: in progress, Ludovic Mouline, A modeling Framework for context representation, 2015, J. Bourcier, O. Barais, F. Fouquet, Y. Le Traon
- PhD: in progress, Erwan Picard, Legal aspect of online profiling, 2015, M.Boizard, B. Baudry

### 10.2.3. Juries

### 10.2.3.1. Benoit Baudry

was in the examination committee of the following PhD thesis:

- Pierre Laperdrix, September 2017, Univ Rennes I, Supervisor
- Marcelino Rodriguez-Cancio, December 2017, Univ Rennes I, Supervisor
- Xavier Devroey, June 2017, Univ Namur, Referee

### 10.2.3.2. Olivier Barais

was in the examination committee of the following HDR:

- Mohammed Boussa, 2017, Univ RennesI, Supervisor
- hamza ouarnoughi, 2017, Univ Brest, Reviewer
- Ternava Xhevahire, 2017, Univ Nice, Reviewer
- Colin AYGALINC, 2017, Univ Grenoble, Reviewer
- Fadwa Rekik, May 2017, Univ Paris Saclay, Reviewer

### 10.2.3.3. Benoit Combemale

was in the examination committee of the following PhD thesis:

- Ulyana Tikhonova, November 2017, TU/e, The Netherlands, Reviewer
- Cyril Cecchinel, November 2017, Univ. Nice, Examiner
- Rodriguez Cancio, December 2017, Univ Rennes 1, Supervisor

### **10.3.** Popularization

- IT industry forums . Pierre Laperdrix presented his work on the mitigation of browser fingerprinting at the Mozilla Festival <sup>0</sup>. The results of STAMP <sup>0</sup> have been presented at EclipseCon <sup>0</sup>, the Paris Open Source Summit <sup>0</sup>, OW2' con <sup>0</sup> and the cloud computing world expo <sup>0</sup>.
- Science festivals . Jean-Marc Jézéquel and Benoit Baudry have presented overviews of research issues in Cybersecurity to general public audiences in several events, such as the 2017 Pint of Science <sup>0</sup> and the science en bobine festival <sup>0</sup>.

# 11. Bibliography

### Major publications by the team in recent years

 B. BAUDRY, M. MONPERRUS. The Multiple Facets of Software Diversity: Recent Developments in Year 2000 and Beyond, in "ACM Computing Surveys", 2015, vol. 48, n<sup>o</sup> 1, p. 16:1–16:26, https://hal.inria.fr/hal-01182103.

<sup>0</sup>https://www.stamp-project.eu

- <sup>0</sup>http://www.opensourcesummit.paris/
- <sup>0</sup>https://www.ow2con.org

<sup>&</sup>lt;sup>0</sup>https://mozillafestival.org/

<sup>&</sup>lt;sup>0</sup>https://www.eclipsecon.org/europe2017/

<sup>&</sup>lt;sup>0</sup>http://www.cloudcomputing-world.com/

<sup>&</sup>lt;sup>0</sup>https://pintofscience.fr/

<sup>&</sup>lt;sup>0</sup>https://sciencesenbobines.org/rennes-35/

- [2] A. BLOUIN, N. MOHA, B. BAUDRY, H. SAHRAOUI, J.-M. JÉZÉQUEL. Assessing the Use of Slicing-based Visualizing Techniques on the Understanding of Large Metamodels, in "Information and Software Technology", 2015, vol. 62, p. 124 - 142 [DOI: 10.1016/J.INFSOF.2015.02.007], https://hal.inria.fr/hal-01120558.
- [3] M. BOUSSAA, O. BARAIS, B. BAUDRY, G. SUNYÉ.NOTICE: A Framework for Non-functional Testing of Compilers, in "Proc. of the Int. Conf. on Software Quality, Reliability & Security (QRS)", August 2016, https://hal.archives-ouvertes.fr/hal-01344835.
- [4] G. BÉCAN, M. ACHER, B. BAUDRY, S. BEN NASR.Breathing Ontological Knowledge Into Feature Model Synthesis: An Empirical Study, in "Empirical Software Engineering", 2015, vol. 21, n<sup>o</sup> 4, p. 1794–1841 [DOI: 10.1007/s10664-014-9357-1], https://hal.inria.fr/hal-01096969.
- [5] G. BÉCAN, N. SANNIER, M. ACHER, O. BARAIS, A. BLOUIN, B. BAUDRY. Automating the Formalization of Product Comparison Matrices, in "Proc. of the Int. Conf. on Automated Software Engineering (ASE)", September 2014 [DOI: 10.1145/2642937.2643000], https://hal.inria.fr/hal-01058440.
- [6] B. COMBEMALE, J. DEANTONI, B. BAUDRY, R. B. FRANCE, J.-M. JÉZÉQUEL, J. GRAY. Globalizing Modeling Languages, in "IEEE Computer", June 2014, p. 10-13, https://hal.inria.fr/hal-00994551.
- [7] B. COMBEMALE, J. DEANTONI, M. E. VARA LARSEN, F. MALLET, O. BARAIS, B. BAUDRY, R. FRANCE. *Reifying Concurrency for Executable Metamodeling*, in "Proc. of the Int. Conf. on Software Language Engineering", October 2013, p. 365-384 [DOI: 10.1007/978-3-319-02654-1\_20], https://hal.inria.fr/hal-00850770.
- [8] J.-M. DAVRIL, E. DELFOSSE, N. HARIRI, M. ACHER, J. CLELANG-HUANG, P. HEYMANS. Feature Model Extraction from Large Collections of Informal Product Descriptions, in "Proc. of the Europ. Software Engineering Conf. and the ACM SIGSOFT Symp. on the Foundations of Software Engineering (ESEC/FSE)", September 2013, p. 290-300 [DOI: 10.1145/2491411.2491455], https://hal.inria.fr/hal-00859475.
- [9] T. DEGUEULE, B. COMBEMALE, A. BLOUIN, O. BARAIS, J.-M. JÉZÉQUEL. Melange: A Meta-language for Modular and Reusable Development of DSLs, in "Proc. of the Int. Conf. on Software Language Engineering (SLE)", October 2015, https://hal.inria.fr/hal-01197038.
- [10] J. A. GALINDO, M. ALFÉREZ, M. ACHER, B. BAUDRY, D. BENAVIDES. A Variability-Based Testing Approach for Synthesizing Video Sequences, in "Proc. of the Int. Symp. on Software Testing and Analysis (ISSTA)", July 2014, https://hal.inria.fr/hal-01003148.
- [11] I. GONZALEZ-HERRERA, J. BOURCIER, E. DAUBERT, W. RUDAMETKIN, O. BARAIS, F. FOUQUET, J.-M. JÉZÉQUEL, B. BAUDRY. ScapeGoat: Spotting abnormal resource usage in component-based reconfigurable software systems, in "Journal of Systems and Software", 2016 [DOI: 10.1016/J.JSS.2016.02.027], https://hal.inria.fr/hal-01354999.
- [12] J.-M. JÉZÉQUEL, B. COMBEMALE, O. BARAIS, M. MONPERRUS, F. FOUQUET. Mashup of Meta-Languages and its Implementation in the Kermeta Language Workbench, in "Software and Systems Modeling", 2015, vol. 14, n<sup>o</sup> 2, p. 905-920, https://hal.inria.fr/hal-00829839.
- [13] P. LAPERDRIX, W. RUDAMETKIN, B. BAUDRY. Beauty and the Beast: Diverting modern web browsers to build unique browser fingerprints, in "Proc. of the Symp. on Security and Privacy (S&P)", May 2016, https:// hal.inria.fr/hal-01285470.

- [14] M. RODRIGUEZ-CANCIO, B. COMBEMALE, B. BAUDRY. Automatic Microbenchmark Generation to Prevent Dead Code Elimination and Constant Folding, in "Proc. of the Int. Conf. on Automated Software Engineering (ASE)", September 2016, https://hal.inria.fr/hal-01343818.
- [15] M. TRICOIRE, O. BARAIS, M. LEDUC, J. BOURCIER, F. FOUQUET, G. NAIN, L. MOULINE, G. SUNYÉ, B. MORIN. *KevoreeJS: Enabling Dynamic Software Reconfigurations in the Browser*, in "Proc. of WICSA and CompArch", April 2016 [DOI : 10.1109/CBSE.2016.20], https://hal.inria.fr/hal-01354997.

### **Publications of the year**

### **Doctoral Dissertations and Habilitation Theses**

- [16] M. BOUSSAA. *Automatic non-functional testing and tuning of configurable generators*, Université Rennes 1, September 2017, https://tel.archives-ouvertes.fr/tel-01598821.
- [17] P. LAPERDRIX.Browser Fingerprinting: Exploring Device Diversity to Augment Authentication and Build Client-Side Countermeasures, INSA Rennes, October 2017, https://hal.inria.fr/tel-01621257.

### **Articles in International Peer-Reviewed Journal**

- [18] M. ACHER, R. E. LOPEZ-HERREJON, R. RABISER. Teaching Software Product Lines: A Snapshot of Current Practices and Challenges, in "ACM Transactions of Computing Education", May 2017, https://hal.inria.fr/hal-01522779.
- [19] M. ALFÉREZ, M. ACHER, J. A. GALINDO, B. BAUDRY, D. BENAVIDES. Modeling Variability in the Video Domain: Language and Experience Report, in "Software Quality Journal", January 2018, p. 1-28, https://hal. inria.fr/hal-01688247.
- [20] S. BEN NASR, G. BÉCAN, M. ACHER, J. F. F. BOSCO, N. SANNIER, B. BAUDRY, J.-M. DAVRIL. Automated Extraction of Product Comparison Matrices From Informal Product Descriptions, in "Journal of Systems and Software", 2017, vol. 124, p. 82 - 103 [DOI: 10.1016/J.JSS.2016.11.018], https://hal.inria.fr/hal-01427218.
- [21] E. BOUSSE, D. LEROY, B. COMBEMALE, M. WIMMER, B. BAUDRY. Omniscient Debugging for Executable DSLs, in "Journal of Systems and Software", November 2017, vol. 137, p. 261-288 [DOI: 10.1016/J.JSS.2017.11.025], https://hal.inria.fr/hal-01662336.
- [22] E. BOUSSE, T. MAYERHOFER, B. COMBEMALE, B. BAUDRY.Advanced and efficient execution trace management for executable domain-specific modeling languages, in "Software and Systems Modeling", May 2017, p. 1–37 [DOI: 10.1007/s10270-017-0598-5], https://hal.inria.fr/hal-01614377.
- [23] K. CORRE, O. BARAIS, G. SUNYÉ, V. FREY, J.-M. CROM. Why can't users choose their identity providers on the web?, in "Proceedings on Privacy Enhancing Technologies", January 2017, vol. 2017, n<sup>o</sup> 3, p. 72-86 [DOI: 10.1515/POPETS-2017-0029], https://hal.archives-ouvertes.fr/hal-01611048.
- [24] K. CORRE, S. BÉCOT, O. BARAIS, G. SUNYÉ. A WebRTC Extension to Allow Identity Negotiation at Runtime, in "Lecture Note in Computer Science", 2017, vol. 10360, p. 412-419 [DOI: 10.1007/978-3-319-60131-1\_27], https://hal.archives-ouvertes.fr/hal-01611057.

- [25] B. DANGLOT, P. PREUX, B. BAUDRY, M. MONPERRUS. Correctness Attraction: A Study of Stability of Software Behavior Under Runtime Perturbation, in "Empirical Software Engineering", 2017, https://arxiv. org/abs/1611.09187 [DOI: 10.1007/s10664-017-9571-8], https://hal.archives-ouvertes.fr/hal-01378523.
- [26] D. A. MÉNDEZ-ACUÑA, J. A. GALINDO, B. COMBEMALE, A. BLOUIN, B. BAUDRY. Reverse Engineering Language Product Lines from Existing DSL Variants, in "Journal of Systems and Software", May 2017 [DOI: 10.1016/J.JSS.2017.05.042], https://hal.inria.fr/hal-01524632.
- [27] P. TEMPLE, M. ACHER, J.-M. JEZEQUEL, O. BARAIS.Learning-Contextual Variability Models, in "IEEE Software", November 2017, vol. 34, n<sup>o</sup> 6, p. 64-70 [DOI: 10.1109/MS.2017.4121211], https://hal.inria.fr/ hal-01659137.

### **International Conferences with Proceedings**

- [28] M. ACHER, P. TEMPLE, J.-M. JEZEQUEL, J. Á. GALINDO DUARTE, J. MARTINEZ, T. ZIADI. VaryLaTeX: Learning Paper Variants That Meet Constraints, in "VaMoS 2018 - 12th International Workshop on Variability Modelling of Software-Intensive Systems", Madrid, Spain, February 2018, p. 1-6, https://hal.inria.fr/hal-01659161.
- [29] O. BARAIS, B. COMBEMALE, A. WORTMANN. Language Engineering with the GEMOC Studio, in "ICSAW 2017 - IEEE International Conference on Software Architecture Workshops", Gothenburg, Sweden, IEEE (editor), April 2017, 3, https://hal.inria.fr/hal-01609576.
- [30] A. BENELALLAM, T. HARTMANN, L. MOULINE, F. FOUQUET, J. BOURCIER, O. BARAIS, Y. LE TRAON. *Raising Time Awareness in Model-Driven Engineering*, in "ACM/IEEE 20th International Conference on Model Driven Engineering Languages and Systems", Austin, Texas, United States, September 2017, https:// hal.archives-ouvertes.fr/hal-01580554.
- [31] A. GÓMEZ-BOIX, P. LAPERDRIX, B. BAUDRY. Fingerprinting mobile devices: A short analysis, in "CIEL 2017 - 6ème Conférence en IngénieriE du Logiciel", Montpellier, France, June 2017, https://hal.inria.fr/hal-01611101.
- [32] A. HALIN, A. NUTTINCK, M. ACHER, X. DEVROEY, G. PERROUIN, P. HEYMANS. Yo Variability! JHipster: A Playground for Web-Apps Analyses, in "11th International Workshop on Variability Modelling of Softwareintensive Systems", Eindhoven, Netherlands, February 2017, p. 44 - 51 [DOI: 10.1145/3023956.3023963], https://hal.inria.fr/hal-01468084.
- [33] P. LAPERDRIX, B. BAUDRY, V. MISHRA. FPRandom: Randomizing core browser objects to break advanced device fingerprinting techniques, in "ESSoS 2017 - 9th International Symposium on Engineering Secure Software and Systems", Bonn, Germany, July 2017, 17, https://hal.inria.fr/hal-01527580.
- [34] G. LE MOULEC, F. ARGELAGUET, V. GOURANTON, A. BLOUIN, B. ARNALDLAGENT: Automatic Generation of Experimental Protocol Runtime, in "ACM Symposium on Virtual Reality Software and Technology (VRST)", Gothenburg, Sweden, Virtual Reality Software and Technology, November 2017, https://hal. archives-ouvertes.fr/hal-01613873.
- [35] M. LEDUC, T. DEGUEULE, B. COMBEMALE, T. VAN DER STORM, O. BARAIS. Revisiting Visitors for Modular Extension of Executable DSMLs, in "ACM/IEEE 20th International Conference on Model Driven Engineering Languages and Systems", Austin, United States, September 2017, https://hal.inria.fr/hal-01568169.

- [36] D. LEROY, E. BOUSSE, M. WIMMER, B. COMBEMALE, W. SCHWINGER. Create and Play your Pac-Man Game with the GEMOC Studio (Tool Demonstration), in "EXE 2017 - 3rd International Workshop on Executable Modeling", Austin, United States, September 2017, p. 1-6, https://hal.inria.fr/hal-01651801.
- [37] Y. NDIAYE, N. AILLERY, O. BARAIS, A. BLOUIN, A. BOUABDALLAH.*Modélisation et Évaluation de la Sécurité des IHM*, in "CIEL 2017 : 6ème Conférence en IngénieriE du Logiciel", Montpellier, France, June 2017, https://hal.inria.fr/hal-01611324.
- [38] Q. PLAZAR, M. ACHER, S. BARDIN, A. GOTLIEB. *Efficient and Complete FD-Solving for Extended Array Constraints* \*, in "IJCAI 2017", Melbourne, Australia, August 2017, https://hal.archives-ouvertes.fr/hal-01545557.
- [39] A. VASTEL, P. LAPERDRIX, W. RUDAMETKIN, R. ROUVOY.FP-STALKER: Tracking Browser Fingerprint Evolutions, in "IEEE S&P 2018 - 39th IEEE Symposium on Security and Privacy", San Francisco, United States, B. PARNO, C. KRUEGEL (editors), Proceedings of the 39th IEEE Symposium on Security and Privacy (S&P), IEEE, May 2018, p. 1-14, https://hal.inria.fr/hal-01652021.

### Scientific Books (or Scientific Book chapters)

[40] T. DEGUEULE, B. COMBEMALE, J.-M. JÉZÉQUEL. On Language Interfaces, in "PAUSE: Present And Ulterior Software Engineering", B. MEYER, M. MAZZARA (editors), Springer, February 2017, https://hal. inria.fr/hal-01424909.

### **Books or Proceedings Editing**

- [41] J. PELAY, F. GUILLEMIN, O. BARAIS (editors). Verifying the configuration of Virtualized Network Functions in Software Defined Networks, b<>com, 2017, p. 1-6, Accepted version of the paper but with two formulas which had to be deleted in the final version for size reasons, https://hal.archives-ouvertes.fr/hal-01657866.
- [42] S. VAN MIERLO, E. BOUSSE, H. VANGHELUWE, M. WIMMER, C. VERBRUGGE, M. GOGOLLA, M. TICHY, A. BLOUIN (editors). Report on the 1 st International Workshop on Debugging in Model-Driven Engineering (MDEbug'17), December 2017, p. 1-6, https://hal.inria.fr/hal-01665572.

### **Research Reports**

- [43] P. TEMPLE, M. ACHER, J.-M. JÉZÉQUEL, L. NOEL-BARON, J. A. GALINDO. Learning-Based Performance Specialization of Configurable Systems, IRISA, Inria Rennes; University of Rennes 1, February 2017, https:// hal.archives-ouvertes.fr/hal-01467299.
- [44] A. WORTMANN, B. COMBEMALE, O. BARAIS. A Systematic Mapping Study on Modeling for Industry 4.0, Inria Rennes - Bretagne Atlantique and University of Rennes 1, France, April 2017, n<sup>o</sup> RR-9062, 25, https:// hal.inria.fr/hal-01514421.

### **Scientific Popularization**

[45] J.-M. JÉZÉQUEL, J. JONGWANE. Comment maîtriser la complexité des logiciels ?, in "Interstices", February 2017, https://hal.inria.fr/hal-01503821.

[46] L. MOULINE, T. HARTMANN, F. FOUQUET, Y. LE TRAON, J. BOURCIER, O. BARAIS. Weaving Rules into Models@run.time for Embedded Smart Systems, in "Programming 2017 - Companion to the first International Conference on the Art, Science and Engineering of Programming", Brussels, Belgium, April 2017, p. 1 - 6 [DOI: 10.1145/3079368.3079394], https://hal.inria.fr/hal-01609796.

### **Other Publications**

- [47] A. BLOUIN, V. LELLI, B. BAUDRY, F. COULON. User Interface Design Smell: Automatic Detection and Refactoring of Blob Listeners, April 2017, working paper or preprint, https://hal.inria.fr/hal-01499106.
- [48] G. LE MOULEC, A. BLOUIN, V. GOURANTON, B. ARNALDI. *Automatic Production of End User Documentation for DSLs*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01549042.
- [49] M. MONPERRUS, B. DANGLOT, O. VERA-PEREZ, Z. YU, B. BAUDRY. *The Emerging Field of Test Amplification: A Survey*, November 2017, https://arxiv.org/abs/1705.10692 working paper or preprint, https://hal.archives-ouvertes.fr/hal-01634288.

### **References in notes**

- [50] A. ARCURI, L. C. BRIAND. A practical guide for using statistical tests to assess randomized algorithms in software engineering, in "ICSE", 2011, p. 1-10.
- [51] A. AVIZIENIS. *The N-version approach to fault-tolerant software*, in "Software Engineering, IEEE Transactions on", 1985, n<sup>o</sup> 12, p. 1491–1501.
- [52] F. BACHMANN, L. BASS. Managing variability in software architectures, in "SIGSOFT Softw. Eng. Notes", 2001, vol. 26, n<sup>o</sup> 3, p. 126–132.
- [53] F. BALARIN, Y. WATANABE, H. HSIEH, L. LAVAGNO, C. PASSERONE, A. SANGIOVANNI-VINCENTELLI. *Metropolis: An integrated electronic system design environment*, in "Computer", 2003, vol. 36, n<sup>o</sup> 4, p. 45–52.
- [54] E. BANIASSAD, S. CLARKE. Theme: an approach for aspect-oriented analysis and design, in "26th International Conference on Software Engineering (ICSE)", 2004, p. 158-167.
- [55] E. G. BARRANTES, D. H. ACKLEY, S. FORREST, D. STEFANOVIĆ.*Randomized instruction set emulation*, in "ACM Transactions on Information and System Security (TISSEC)", 2005, vol. 8, n<sup>o</sup> 1, p. 3–40.
- [56] D. BATORY, R. E. LOPEZ-HERREJON, J.-P. MARTIN. Generating Product-Lines of Product-Families, in "ASE '02: Automated software engineering", IEEE, 2002, p. 81–92.
- [57] S. BECKER, H. KOZIOLEK, R. REUSSNER. The Palladio component model for model-driven performance prediction, in "Journal of Systems and Software", January 2009, vol. 82, n<sup>o</sup> 1, p. 3–22.
- [58] N. BENCOMO. On the use of software models during software execution, in "MISE '09: Proceedings of the 2009 ICSE Workshop on Modeling in Software Engineering", IEEE Computer Society, May 2009.
- [59] A. BEUGNARD, J.-M. JÉZÉQUEL, N. PLOUZEAU. Contract Aware Components, 10 years after, in "WCSI", 2010, p. 1-11.

- [60] J. BOSCH. Design and use of software architectures: adopting and evolving a product-line approach, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 2000.
- [61] J. BOSCH, G. FLORIJN, D. GREEFHORST, J. KUUSELA, J. H. OBBINK, K. POHL. Variability Issues in Software Product Lines, in "PFE '01: Revised Papers from the 4th International Workshop on Software Product-Family Engineering", London, UK, Springer-Verlag, 2002, p. 13–21.
- [62] L. C. BRIAND, E. ARISHOLM, S. COUNSELL, F. HOUDEK, P. THÉVENOD-FOSSE. *Empirical studies of object-oriented artifacts, methods, and processes: state of the art and future directions*, in "Empirical Software Engineering", 1999, vol. 4, n<sup>o</sup> 4, p. 387–404.
- [63] J. T. BUCK, S. HA, E. A. LEE, D. G. MESSERSCHMITT. *Ptolemy: A framework for simulating and prototyping heterogeneous systems*, in "Int. Journal of Computer Simulation", 1994.
- [64] T. BURES, P. HNETYNKA, F. PLASIL. Sofa 2.0: Balancing advanced features in a hierarchical component model, in "Software Engineering Research, Management and Applications, 2006. Fourth International Conference on", IEEE, 2006, p. 40–48.
- [65] B. H. C. CHENG, R. LEMOS, H. GIESE, P. INVERARDI, J. MAGEE, J. ANDERSSON, B. BECKER, N. BENCOMO, Y. BRUN, B. CUKIC, G. MARZO SERUGENDO, S. DUSTDAR, A. FINKELSTEIN, C. GACEK, K. GEIHS, V. GRASSI, G. KARSAI, H. M. KIENLE, J. KRAMER, M. LITOIU, S. MALEK, R. MIRANDOLA, H. A. MÜLLER, S. PARK, M. SHAW, M. TICHY, M. TIVOLI, D. WEYNS, J. WHITTLE., D. HUTCHISON, T. KANADE, J. KITTLER, J. M. KLEINBERG, F. MATTERN, J. C. MITCHELL, M. NAOR, O. NIERSTRASZ, C. PANDU RANGAN, B. STEFFEN, M. SUDAN, D. TERZOPOULOS, D. TYGAR, M. Y. VARDI, G. WEIKUM, B. H. C. CHENG, R. LEMOS, H. GIESE, P. INVERARDI, J. MAGEE (editors)Software Engineering for Self-Adaptive Systems: A Research Roadmap, Betty H. C. Cheng, Rogério de Lemos, Holger Giese, Paola Inverardi, and Jeff Magee, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, vol. 5525.
- [66] J. COPLIEN, D. HOFFMAN, D. WEISS. Commonality and Variability in Software Engineering, in "IEEE Software", 1998, vol. 15, n<sup>o</sup> 6, p. 37–45.
- [67] I. CRNKOVIC, S. SENTILLES, A. VULGARAKIS, M. R. CHAUDRON. A classification framework for software component models, in "Software Engineering, IEEE Transactions on", 2011, vol. 37, n<sup>o</sup> 5, p. 593–615.
- [68] K. CZARNECKI, U. W. EISENECKER. *Generative programming: methods, tools, and applications*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 2000.
- [69] R. DEMILLI, A. J. OFFUTT. *Constraint-based automatic test data generation*, in "Software Engineering, IEEE Transactions on", 1991, vol. 17, n<sup>o</sup> 9, p. 900–910.
- [70] K. DEB, A. PRATAP, S. AGARWAL, T. MEYARIVAN. *A fast and elitist multiobjective genetic algorithm: NSGA-II*, in "Evolutionary Computation, IEEE Transactions on", 2002, vol. 6, n<sup>o</sup> 2, p. 182–197.
- [71] S. FORREST, A. SOMAYAJI, D. H. ACKLEY. *Building diverse computer systems*, in "Operating Systems, 1997., The Sixth Workshop on Hot Topics in", IEEE, 1997, p. 67–72.
- [72] R. B. FRANCE, B. RUMPE. Model-driven Development of Complex Software: A Research Roadmap, in "Proceedings of the Future of Software Engineering Symposium (FOSE '07)", L. C. BRIAND, A. L. WOLF (editors), IEEE, 2007, p. 37–54.

- [73] S. FREY, F. FITTKAU, W. HASSELBRING. Search-based genetic optimization for deployment and reconfiguration of software in the cloud, in "Proceedings of the 2013 International Conference on Software Engineering", IEEE Press, 2013, p. 512–521.
- [74] G. HALMANS, K. POHL. Communicating the Variability of a Software-Product Family to Customers, in "Software and System Modeling", 2003, vol. 2, n<sup>o</sup> 1, p. 15-36.
- [75] C. HARDEBOLLE, F. BOULANGER. ModHel'X: A component-oriented approach to multi-formalism modeling, in "Models in Software Engineering", Springer, 2008, p. 247–258.
- [76] M. HARMAN, B. F. JONES. Search-based software engineering, in "Information and Software Technology", 2001, vol. 43, n<sup>o</sup> 14, p. 833–839.
- [77] H. HEMMATI, L. C. BRIAND, A. ARCURI, S. ALI.An enhanced test case selection approach for model-based testing: an industrial case study, in "SIGSOFT FSE", 2010, p. 267-276.
- [78] J. HUTCHINSON, J. WHITTLE, M. ROUNCEFIELD, S. KRISTOFFERSEN. *Empirical assessment of MDE in industry*, in "Proceedings of the 33rd International Conference on Software Engineering (ICSE '11)", R. N. TAYLOR, H. GALL, N. MEDVIDOVIC (editors), ACM, 2011, p. 471–480.
- [79] J.-M. JÉZÉQUEL. Model Driven Design and Aspect Weaving, in "Journal of Software and Systems Modeling (SoSyM)", may 2008, vol. 7, n<sup>o</sup> 2, p. 209–218, http://www.irisa.fr/triskell/publis/2008/Jezequel08a.pdf.
- [80] K. C. KANG, S. G. COHEN, J. A. HESS, W. E. NOVAK, A. S. PETERSON. Feature-Oriented Domain Analysis (FODA) Feasibility Study, Carnegie-Mellon University Software Engineering Institute, November 1990.
- [81] J. KRAMER, J. MAGEE.Self-Managed Systems: an Architectural Challenge, in "Future of Software Engineering", IEEE, 2007, p. 259–268.
- [82] K.-K. LAU, P. V. ELIZONDO, Z. WANG. Exogenous connectors for software components, in "Component-Based Software Engineering", Springer, 2005, p. 90–106.
- [83] P. MCMINN.Search-based software test data generation: a survey, in "Software Testing, Verification and Reliability", 2004, vol. 14, n<sup>o</sup> 2, p. 105–156.
- [84] J. MEEKEL, T. B. HORTON, C. MELLONE. Architecting for Domain Variability, in "ESPRIT ARES Workshop", 1998, p. 205-213.
- [85] A. M. MEMON.An event-flow model of GUI-based applications for testing, in "Software Testing, Verification and Reliability", 2007, vol. 17, n<sup>o</sup> 3, p. 137–157.
- [86] B. MORIN, O. BARAIS, J.-M. JÉZÉQUEL, F. FLEUREY, A. SOLBERG. Models at Runtime to Support Dynamic Adaptation, in "IEEE Computer", October 2009, p. 46-53, http://www.irisa.fr/triskell/publis/2009/ Morin09f.pdf.
- [87] P.-A. MULLER, F. FLEUREY, J.-M. JÉZÉQUEL. Weaving Executability into Object-Oriented Meta-Languages, in "Proc. of MODELS/UML'2005", Jamaica, LNCS, Springer, 2005.

- [88] R. MÉLISSON, P. MERLE, D. ROMERO, R. ROUVOY, L. SEINTURIER. Reconfigurable run-time support for distributed service component architectures, in "the IEEE/ACM international conference", New York, New York, USA, ACM Press, 2010, 171.
- [89] C. NEBUT, Y. LE TRAON, J.-M. JÉZÉQUEL.System Testing of Product Families: from Requirements to Test Cases, Springer Verlag, 2006, p. 447–478, http://www.irisa.fr/triskell/publis/2006/Nebut06b.pdf.
- [90] C. NEBUT, S. PICKIN, Y. LE TRAON, J.-M. JÉZÉQUEL. Automated Requirements-based Generation of Test Cases for Product Families, in "Proc. of the 18th IEEE International Conference on Automated Software Engineering (ASE'03)", 2003, http://www.irisa.fr/triskell/publis/2003/nebut03b.pdf.
- [91] L. M. NORTHROP.SEI's Software Product Line Tenets, in "IEEE Softw.", 2002, vol. 19, nº 4, p. 32-40.
- [92] L. M. NORTHROP.A Framework for Software Product Line Practice, in "Proceedings of the Workshop on Object-Oriented Technology", Springer-Verlag London, UK, 1999, p. 365–376.
- [93] I. OBER, S. GRAF, I. OBER. Validating timed UML models by simulation and verification, in "International Journal on Software Tools for Technology Transfer", 2006, vol. 8, n<sup>o</sup> 2, p. 128–145.
- [94] D. L. PARNAS. On the Design and Development of Program Families, in "IEEE Trans. Softw. Eng.", 1976, vol. 2, n<sup>o</sup> 1, p. 1–9.
- [95] S. PICKIN, C. JARD, T. JÉRON, J.-M. JÉZÉQUEL, Y. LE TRAON. Test Synthesis from UML Models of Distributed Software, in "IEEE Transactions on Software Engineering", April 2007, vol. 33, nº 4, p. 252–268.
- [96] K. POHL, G. BÖCKLE, F. J. VAN DER LINDEN. Software Product Line Engineering: Foundations, Principles and Techniques, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.
- [97] B. RANDELL.System structure for software fault tolerance, in "Software Engineering, IEEE Transactions on", 1975, nº 2, p. 220–232.
- [98] M. RINARD. *Obtaining and reasoning about good enough software*, in "Proceedings of Annual Design Automation Conference (DAC)", 2012, p. 930-935.
- [99] J. ROTHENBERG, L. E. WIDMAN, K. A. LOPARO, N. R. NIELSEN. *The Nature of Modeling*, in "in Artificial Intelligence, Simulation and Modeling", John Wiley & Sons, 1989, p. 75–92.
- [100] P. RUNESON, M. HÖST. *Guidelines for conducting and reporting case study research in software engineering*, in "Empirical Software Engineering", 2009, vol. 14, n<sup>o</sup> 2, p. 131–164.
- [101] D. SCHMIDT.Guest Editor's Introduction: Model-Driven Engineering, in "IEEE Computer", 2006, vol. 39, n<sup>o</sup> 2, p. 25–31.
- [102] F. SHULL, J. SINGER, D. I. SJBERG. Guide to advanced empirical software engineering, Springer, 2008.
- [103] S. SIDIROGLOU-DOUSKOS, S. MISAILOVIC, H. HOFFMANN, M. RINARD. Managing performance vs. accuracy trade-offs with loop perforation, in "Proc. of the Symp. on Foundations of software engineering", New York, NY, USA, ESEC/FSE '11, ACM, 2011, p. 124-134.

- [104] J. STEEL, J.-M. JÉZÉQUEL. On Model Typing, in "Journal of Software and Systems Modeling (SoSyM)", December 2007, vol. 6, n<sup>o</sup> 4, p. 401–414, http://www.irisa.fr/triskell/publis/2007/Steel07a.pdf.
- [105] C. SZYPERSKI, D. GRUNTZ, S. MURER. Component software: beyond object-oriented programming, Addison-Wesley, 2002.
- [106] J.-C. TRIGAUX, P. HEYMANS. Modelling variability requirements in Software Product Lines: a comparative survey, FUNDP Namur, 2003.
- [107] M. UTTING, B. LEGEARD. Practical model-based testing: a tools approach, Morgan Kaufmann, 2010.
- [108] P. VROMANT, D. WEYNS, S. MALEK, J. ANDERSSON. On interacting control loops in self-adaptive systems, in "SEAMS 2011", ACM, 2011, p. 202–207.
- [109] C. YILMAZ, M. B. COHEN, A. A. PORTER. *Covering arrays for efficient fault characterization in complex configuration spaces*, in "Software Engineering, IEEE Transactions on", 2006, vol. 32, n<sup>o</sup> 1, p. 20–34.
- [110] Z. A. ZHU, S. MISAILOVIC, J. A. KELNER, M. RINARD. Randomized accuracy-aware program transformations for efficient approximate computations, in "Proc. of the Symp. on Principles of Programming Languages (POPL)", 2012, p. 441-454.
- [111] T. ZIADI, J.-M. JÉZÉQUEL. Product Line Engineering with the UML: Deriving Products, Springer Verlag, 2006, p. 557-586.

# **Project-Team DYLISS**

# Dynamics, Logics and Inference for biological Systems and Sequences

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: CNRS

**Université Rennes 1** 

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Computational Biology

### **Table of contents**

1.	Personnel	269
2.	Overall Objectives	270
3.	Research Program	271
	3.1. Modeling knowledge integration with combinatorial constraints	271
	3.2. Modeling the dynamical response of biological systems with logical and (non)-line	ar con-
	straints	272
	3.3. Modeling sequences with formal grammars	272
	3.4. Symbolic methods for model space exploration: Semantic web for life sciences and	Formal
	Concepts Analysis	273
	3.5. Implementing methods in software and platforms	274
	3.5.1. AskOmics platform	275
	3.5.2. AuReMe workspace	275
	3.5.3. FinGoc-tools	276
	3.5.4. Caspo software	276
	3.5.5. Cadbiom package	277
	3.5.6. Logol software	277
	3.5.7. Protomata-suite	277
4.	Application Domains	278
5.	Highlights of the Year	
6.	New Software and Platforms	279
	6.1. AskOmics	279
	6.2. PADMet-utils	280
	6.3. CADBIOM	280
_	6.4. conquests	280
7.	New Results	280
	7.1. Data integration and pre-processing with semantic-based technologies	280
	7.2. Data and knowledge integration based on combinatorial optimization	281
	7.3. Systems biology	282
0	7.4. Sequence and structure annotation	283
ð. 0	Bilateral Contracts and Grants with Industry	283
9.	0.1 Decional Initiativas	<b>203</b>
	0.1.1 Descional initiative: the Ecosyst project	203
	9.1.1. Regional mitiative, the Ecosyst project	203
	0.1.2. Regional partnership with computer science faboratories in Mantes	204
	9.1.5. Regional partnership in warne blology 0.1.4. Regional partnership in agriculture and environmental sciences	204
	9.1.4. Regional partnership in agriculture and environmental sciences	285
	9.1.5. Regional participant in nearth	285
	9.2. Average $3.2.$ Average $3.2.$	205
	9.2.2. Programs funded by research institutions	205
	9.2.2.1. PEPS PEPS: a platform for supporting studies in pharmaco-epidemiolog	v using
	medico-administrative databases	285
	9.2.2.2. Cancer Plan: TGFSvsBio	286
	9.2.2.3. ANR Samosa	286
	9.2.2.4. ANSES Mecagenotox	286
	9.2.3. Programs funded by Inria	287
	9.2.3.1. ADT Complex-biomarkers and ADT Proof of concept	287
	9.2.3.2. IPL Algae in silico	287
	9.2.3.3. IPL Neuromarkers	287

	9.2.3.4. FederatedQueryScaler (Exploratory Research Action)	287
	9.3. European Initiatives	287
	9.4. International Initiatives	287
	9.5. International Research Visitors	288
	9.5.1. Visits of International Scientists	288
	9.5.2. Visits to International Labs	288
10.	Dissemination	
	10.1. Promoting Scientific Activities	288
	10.1.1. Scientific Events Selection	288
	10.1.1.1. Member of the Conference Program Committees	288
	10.1.1.2. Review	288
	10.1.2. Journal	288
	10.1.2.1. Member of the Editorial Boards	288
	10.1.2.2. Reviewer - Reviewing Activities	288
	10.1.3. Invited Talks	288
	10.1.3.1. French seminars	288
	10.1.3.2. Workshops and meetings	288
	10.1.4. Leadership within the Scientific Community	289
	10.1.5. Scientific Expertise	289
	10.1.5.1. International responsibilities	289
	10.1.5.2. National responsibilities	289
	10.1.5.3. National scientific boards	289
	10.1.5.4. Prospective working groups	289
	10.1.5.5. Local responsibilities	289
	10.2. Teaching - Supervision - Juries	290
	10.2.1. Teaching track responsibilities	290
	10.2.2. Course responsibilities	290
	10.2.3. Teaching	290
	10.2.4. Supervision	291
	10.2.5. Juries	292
	10.2.6. Internships	292
	10.3. Popularization	293
	10.3.1. General audience paper	293
	10.3.2. Science en Cour[t]s	293
11.	Bibliography	

### **Project-Team DYLISS**

*Creation of the Team: 2012 January 01, updated into Project-Team: 2013 July 01* **Keywords:** 

### **Computer Science and Digital Science:**

- A3.1.1. Modeling, representation
- A3.1.2. Data management, quering and storage
- A3.1.7. Open data
- A3.2.3. Inference
- A3.2.4. Semantic Web
- A3.2.5. Ontologies
- A3.3. Data and knowledge analysis
- A7.2. Logic in Computer Science
- A8.1. Discrete mathematics, combinatorics
- A8.2. Optimization
- A9.1. Knowledge
- A9.2. Machine learning
- A9.7. AI algorithmics

### **Other Research Topics and Application Domains:**

- B1.1.2. Molecular biology
- B1.1.3. Cellular biology
- B1.1.9. Bioinformatics
- B1.1.11. Systems biology
- B1.1.14. Microbiology
- B2.2.3. Cancer
- B2.2.5. Immune system diseases

# **1. Personnel**

### **Research Scientists**

Anne Siegel [Team leader, CNRS, Senior Researcher, HDR] François Coste [Inria, Researcher] Jacques Nicolas [Inria, Senior Researcher, HDR]

### **Faculty Members**

Catherine Belleannée [Univ de Rennes I, Associate Professor] Olivier Dameron [Univ de Rennes I, Associate Professor, HDR]

### **External Collaborators**

Jérémie Bourdon [Univ de Nantes, HDR] Marie Chevallier [CNRS] Damien Eveillard [Univ de Nantes] François Moreews [INRA] Denis Tagu [INRA, HDR] Nathalie Théret [INSERM, HDR]

### **Technical Staff**

Meziane Aite [Inria] Xavier Garnier [INRA] Camille Trottier [CNRS, until Sep 2017] Pierre Vignet [CNRS] Jeanne Got [CNRS] Arnaud Belcour [CNRS, from Dec 2017]

### PhD Students

Lucas Bourneuf [Univ de Rennes I] Mael Conan [Univ de Rennes I, from Oct 2017] Jean Coquet [Univ de Rennes I] Victorien Delannée [Univ de Rennes I, until Nov 2017] Clémence Frioux [Inria] Julie Laniau [Inria, until Jul 2017 and CNRS from Sep 2017 until Nov 2017] Marine Louarn [Univ de Rennes I, from Oct 2017] Yann Rivault [Univ de Rennes I] Juliette Talibart [Univ de Rennes I, from Nov 2017] Méline Wery [Univ de Rennes I, from Oct 2017]

### **Post-Doctoral Fellows**

Maxime Folschette [Univ de Rennes I, from Nov 2017] Vijay Ingalalli [CNRS, from Jun 2017]

### Visiting Scientist

Oumarou Abdou Arbi [Université de Maradi, from Sep 2017 to Dec 2017]

#### Administrative Assistant

Marie Le Roic [Univ de Rennes I]

# 2. Overall Objectives

### 2.1. Overall objectives

The research domain of the bioinformatics Dyliss team is sequence analysis and systems biology. Our main goal in biology is to characterize groups of genetic actors that control the phenotypic answer of species when challenged by their environment. The team explores methods in the field of formal systems, more precisely in knowledge representation, constraints programming, multi-scale analysis of dynamical systems, and machine learning. Our goal is to identify key regulators of the environmental response by structuring and reasoning on information which combines physiological responses measured with omics technologies (RNA-seq, metabolomics, proteomics), genetic information from their long-distant cousins and knowledge about regulation and metabolic pathways stored in public repositories.

The main challenges we face are data incompleteness and heterogeneity. We favor the construction and study of a "space of feasible models or hypotheses" including known constraints and facts on a living system *rather than searching for a single optimized model*. We develop methods allowing a precise investigation of this space of hypotheses. Therefore, we are in position of developing experimental strategies to progressively shrink the space of hypotheses and gain in the understanding of the system. Importantly, one should notice that our models span a quite large spectrum of discrete structures: oriented graphs, boolean networks, automata, or expressive grammars.

More concretely, the steps of the analysis are to (*i*) formalize and integrate in a set of logical or grammatical constraints both generic knowledge information (literature-based regulatory pathways, diversity of molecular functions, DNA patterns associated with molecular mechanisms) and species-specific information (physiological response to perturbation, sequencing...); (*ii*) investigate the space of admissible models and exhibit its main features by solving combinatorial optimization problems; (*iii*) identify corresponding genomic products within sequences. At each of these steps, we rely on symbolic methods for model space exploration (ontologies and formal concepts analysis).

We target applications for which large-scale heterogeneous data about a specific but complex physiological phenotype are available. Existing long-term partnerships with biological labs give strong support to this choice. In marine biology, we collaborate closely with the Station biologique de Roscoff (*Idealg*, Investissement avenir "Bioressources et Biotechnologies"). We also collaborate with other teams of Inria in the *IPL Algae In Silico project to understand the metabolism of a micro-algae*. In environmental microbiology we collaborate both with the CRG in Chile in the framework of the Ciric Chilean Inria center (*Ciric-Omics*). In agriculture, our main partners are within the INRA institute in Rennes, with a focus on the understanding of pea-aphids reproduction mode and of breeding animals metabolism (porc, chicken, cow). More recently, we have introduced health as a new application field of the team, especially through the study of large-scale boolean networks and their confrontation with knowledge repositories (collaboration with Inserm, CHU Rennes and Sanofi).

# 3. Research Program

### 3.1. Modeling knowledge integration with combinatorial constraints

Biological networks are built with data-driven approaches aiming at translating genomic information into a functional map. Most methods are based on a probabilistic framework which defines a probability distribution over the set of models. The reconstructed network is then defined as the most likely model given the data.

Our team has investigated an alternative perspective where each data induces a set of constraints – related to the steady state response of the system dynamics – on the set of possible values in a network of fixed topology. The methods that we have developed complete the network with product states at the level of nodes and influence types at the level of edges, able to globally explain experimental data. In other words, the selection of relevant information in the model is no more performed by selecting *the* network with the highest score, but rather by exploring the complete space of models satisfying constraints on the possible dynamics supported by prior knowledge and observations. In the (common) case when there is no model satisfying all the constraints, we relax the problem by introducing new combinatorial optimization problems that introduce the possibility of correcting the data or the knowledge. Common properties to all solutions are considered as a robust information about the system, as they are independent from the choice of a single solution to the optimization problem [77].

Solving these computational issues requires addressing NP-hard qualitative (non-temporal) issues. We have developed a long-term collaboration with Potsdam University in order to use a logical paradigm named **Answer Set Programming** (ASP) [46], [73] to solve these constraint satisfiability and combinatorial optimization issues. Applied on transcriptomic or cancer networks, our methods identified which regions of a large-scale network shall be corrected [47], and proposed robust corrections [6]. This result suggested that this approach was compatible with efficiency, scale and expressivity needed by biological systems.

During the last years, our goal was to provide **formal models of queries on biological networks** with the focus of integrating dynamical information as explicit logical constraints in the modeling process. Using these technologies requires to revisit and reformulate constraint-satisfiability problems at hand in order both to decrease the search space size in the grounding part of the process and to improve the exploration of this search space in the solving part of the process. Concretely, getting logical encoding for the optimization problems forces to clarify the roles and dependencies between parameters involved in the problem. This paves the way to a refinement approach based on a fine investigation of the space of hypotheses in order to make it smaller

and gain in the understanding of the system. Our studies confirmed that logical paradigms are a powerful approach to build and query reconstructed biological systems, in complement to discriminative ("black-box") approaches based on statistical machine-learning. Based on these technologies, we have developed a panel of methods allowing the integration of muli-scale data knowledge, linking genomics, metabolomics, expression data and protein measurement of several phenotypes.

Notice that our main issue is in the field of knowledge representation. More precisely, we do not wish to develop new solvers or grounders, a self-contained computational issue which is addressed by specialized teams such as our collaborator team in Potsdam. Our goal is rather to investigate how the constant progresses in the field of constraint logical programming, shown by the performance of ASP-solvers, are sufficient to address the complexity of constraint-satisfiability and combinatorial optimization issues explored in systems biology. In this direction, we work in close interaction with Potsdam university to feed their research activities whith challenging issues from bioinformatics and, as a feed-back, take benefit of the prototypes they develop.

By exploring the complete space of models, our approach typically produces numerous candidate models compatible with the observations. We began investigating to what extent domain knowledge can further refine the analysis of the set of models by identifying classes of similar models, or by selecting a subset of models that satisfy an additional constraint (for instance, best fit with a set of experiments, or with a minimal size). We anticipate that this will be particularly relevant when studying non-model species for which little is known but valuable information from other species can be transposed or adapted. These efforts consist in developing reasoning methods based on ontologies as formal representation of symbolic knowledge. We use Semantic Web tools such as SPARQL for querying and integrating large sources of external knowledge, and measures of semantic similarity and particularity for analyzing data.

# **3.2.** Modeling the dynamical response of biological systems with logical and (non)-linear constraints

As explained below, Answer Set Programming technologies enable the identification of key controllers based on the integration of static data. As a natural follow-up, we also develop optimization techniques to learn models of the dynamics of a biological system. As before, our strategy is not to select a single model fitting with experimental data but rather to decipher the complete set of families of models which are compatible with the observed response. Our main research line in this field is to decipher the appropriate level of expressivity (in terms of constraints) allowing both to properly report the nature of data and knowledge and to allow for an exhaustive study of the space of feasible models. To implement this strategy, we rely on several constraint programming frameworks, which depend on the model scale and the nature of time-points kinetic measurements. Logical programming (Answer Set Programming) is used to decipher the combinatorics of synchrone Boolean networks explaining static or dynamics response of signaling networks to perturbations (such as measured by phosphoproteomics technologies) [7]. SAT-based approaches are used to decipher the combinatorics of large-scale asynchronous boolean networks. In order to gain in expressivity, we model these networks as guarded-transition network, an extension of Petri nets [40],. Finally, classical learning methods are used to build ad-hoc parameterized numerical models that provide the most parsimonious explanations to experimental measurements.

### **3.3. Modeling sequences with formal grammars**

Once groups of genome products involved in the answer of the species have been identified with integrative or dynamical methods, it remains to characterize the biological actors within genomes. To that goal, we both learn, model and parse formal patterns within DNA, RNA or protein sequences. More precisely, our research on modeling biomolecular sequences with expressive formal grammars focuses on learning such grammars from examples, helping biologists to design their own grammar and providing practical parsing tools.

On the development of **machine learning** algorithms for the induction of grammatical models [66], we have a strong expertise on learning finite state automata. We have proposed an algorithm that learns successfully automata modeling families of (non homologous) functional families of proteins [5], leading to a tool named Protomata-learner. The algorithm is based on a similar fragment merging heuristic approach which reports partial and local alignments contained in a family of sequences. As an example, this tool allowed us to properly model the TNF protein family, a difficult task for classical probabilistic-based approaches. It was also applied successfully to model important enzymatic families of proteins in cyanobacteria [4]. Our future goal is to further demonstrate the relevance of formal language modeling by addressing the question of a fully automatic prediction from the sequence of all the enzymatic families, aiming at improving even more the sensitivity and specificity of the models. As enzyme-substrate interactions are very specific central relations for integrated genome/metabolome studies and are characterized by faint signatures, we shall rely on models for active sites involved in cellular regulation or catalysis mechanisms. This requires to build models gathering both structural and sequence information in order to describe (potentially nested or crossing) long-term dependencies such as contacts of amino-acids that are far in the sequence but close in the 3D protein folding. Our current researches is focused on the inference of Context-Free Grammars including the topological information coming from the structural characterization of active sites.

Using context-free grammars instead of regular patterns increases the complexity of parsing issues. Indeed, efficient parsing tools have been developed to identify patterns within genomes but most of them are restricted to simple regular patterns. Definite Clause Grammars (DCG), a particular form of logical contextfree grammars have been used in various works to model DNA sequence features [86]. An extended formalism, String Variable Grammars (SVGs), introduces variables that can be associated to a string during a pattern search [110], [109]. This increases the expressivity of the formalism towards mildly context sensitive grammars. Thus, those grammars model not only DNA/RNA sequence features but also structural features such as repeats, palindromes, stem/loop or pseudo-knots. A few years ago, we have designed a first tool, STAN (suffix-tree analyser), in order to make it possible to search for a subset of SVG patterns in full chromosome sequences [92]. This tool was used for the recognition of transposable elements in Arabidopsis thaliana [113]. We have enlarged this experience through a new modeling language, called Logol [1]. Generally, a suitable language for the search of particular components in languages has to meet several needs : expressing existing structures in a compact way, using existing databases of motifs, helping the description of interacting components. In other words, the difficulty is to find a good tradeoff between expressivity and complexity to allow the specification of realistic models at genome scale. The Logol language and associated framework have been built in this direction. The Logol specificity besides other SVG-like languages mainly lies in a systematic introduction of constraints on string variables.

# 3.4. Symbolic methods for model space exploration: Semantic web for life sciences and Formal Concepts Analysis

All the methods presented in the previous sections usually result in pools of candidates which equivalently explain the data and knowledge. These candidates can be dynamical systems, compounds, biological sequences, proteins... In any case, the output of our formal methods generally requires *a posteriori* investigation and filtering by domain experts. In order to assist them, we rely on two classes of symbolic technics: Semantic Web technologies and Formal Concept Analysis (FCA). They both aim at the formalization and management of knowledge, that is, the explicitation of relations occuring in structured data. These technics complement each other: the production of relevant concepts in FCA highly depends on the availability of semantic annotations using a controlled set of terms and conversely, building and exploiting ontologies is a complex process that can be made much easier with FCA.

**Integrating heterogenous data with semantic web technologies** The emergence of ontologies in biomedical informatics and bioinformatics happened in parallel with the development of the **Semantic Web** in the computer science community [108]. Let us recall that the Semantic Web is an extension of the current Web that provides an infrastructure integrating data and ontologies in order to support unified reasoning. Since the beginning, life sciences have been a major application domain for the Semantic Web [48]. This was

motivated by the joint evolution of data acquisition capabilities in the biomedical field, and of the methods and infrastructures supporting data analysis (grids, the Internet...), resulting in an explosion of data production in complementary domains [58], [49]. Consequently, Semantic Web technologies have become an integral part of translational medicine and translational bioinformatics [63]. The Linked Open Data project promotes the integration of data sources in machine-processable formats compatible with the Semantic Web [57], with a strong involvement of life sciences in this initiative.

However, a specificity of life sciences "data deluge" is that the proportion of generated data is much higher than in the more general "big data phenomenon", and that these data are highly connected [112]. **The bottleneck that once was data scarcity now lies in the lack of adequate methods supporting data integration, processing and analysis** [88]. Each of these steps typically hinges on domain knowledge, which is why they resist automation. This knowledge can be seen as the set of rules representing in what conditions data can be used or can be combined for inferring new data or new links between data.

In this setting, we are working on the integration of Semantic Web resources with our data analysis methods in order to take existing biological knowledge into account. We have introduced several methods to interpret semantic similarities and particularities [56], [55]. We now focus our attention on the semi-automated construction of RDF abstractions of heterogeneous datasets which can be handled by non-expert users. This allows both to automatically prepare input datasets for the other methods developed in the team and to analyse the output of the methods in a wide knowledge context.

Using Formal concept analysis to explore the results of bioinformatics analyses Formal concept analysis aims at the development of conceptual structures which can be logically activated for the formation of judgments and conclusions [117]. It is used in various domains managing structured data such as knowledge processing, information retrieval or classification [90]. In its most simple form, one considers a binary relation between a set of objects and a set of attributes. In this setting, formal concept analysis formalizes the semantic notions of extension and intension. Concepts are related within a lattice structure (Galois connection) by subconcept-superconcept relations, and this allows drawing causality relations between attribute subsets. In bioinformatics, it has been used to derive phylogenetic relations among groups of organisms [87], a classification task that requires to take into account many-valued Galois connections. We have proposed in a similar way a classification scheme for the problem of protein assignment in a set of protein families [67].

One of the most important issue with concept analysis is due to the fact that current methods remain very sensitive to the presence of uncertainty or incompleteness in data. On the other hand, this apparent defect can be reversed to serve as a marker of incompleteness or inconsistency [68]. Following this inspiration, we have proposed a methodology to tackle the problem of uncertainty on biological networks where edges are mostly predicted links with a high level of false positives [118]. The general idea consists to look for a tradeoff between the simplicity of the conceptual representation and the need to manage exceptions. As a very prospective challenge, we are exploring the idea of using ontologies to help this or to help ontology refinement using concept analysis [93], [52], [102].

More generally, common difficult tasks in this context are visualization, search for local structures (graph mining) and network comparison. Network compression is a good solution for an efficient treatment of all these tasks. This has been used with success in power graphs, which are abstract graphs where nodes are clusters of nodes in the initial graph and edges represent bicliques between two sets of nodes [104]. In fact, concepts are maximal bicliques and we are currently developing the power graph idea in the framework of concept analysis.

### 3.5. Implementing methods in software and platforms

Seven platforms have been developped in the team for the last five years: Askomics, AuReMe, FinGoc, Caspo, Cadbiom, Logol, Protomata. Indeed, one of the team's goals is to facilitate interplays between the tools for biological data analysis and integration. Improvements and novelties of these platforms are described in the "software" section. Our platforms aim at guiding the user to progressively reduce the space of models (families of sequences of genes or proteins, families of keys actors involved in a system response, dynamical models) which are compatible with both knowledge and experimental observations.

Most of our platforms are developed with the support of the GenOuest resource and data center hosted in the IRISA laboratory, including their computer facilities [more info]. It worths considering them into larger dedicated environments to benefit from the expertise of other research groups. The **BioShadock** repository of the GenOuest platform allows one to share the different docker containers that we are developing [website]. The **GenOuest galaxy portal** of the GenOuest platform now provides access to most tools for integrative biology and sequence annotation (access on demand).

### 3.5.1. AskOmics platform

**Goal** Integration and interrogation software for linked biological data based on semantic web technologies [url].

**Description***AskOmics* aims at bridging the gap between end user data and the Linked (Open) Data cloud. It allows heterogeneous bioinformatics data (formatted as tabular files or directly in RDF) to be loaded into a Triple Store system using a user-friendly web interface. *AskOmics* also provides an intuitive graph-based user interface supporting the creation of complex queries that currently require hours of manual searches across tens of spreadsheet files. The elements of interest selected in the graph are then automatically converted into a SPARQL query that is executed on the users's data.

**Originality** Our experience is that end users (i) do not benefit for all the information available in the LOD cloud repositories by lack of SPARQL expertise (understandably: they are biologists and most of them do not have an interest in either learning SPARQL nor in learning how to integrate data); (ii) do not contribute their data back to the LOD cloud. Again, they do not have the expertise nor the resources to produce and maintain datasets and the associated metadata as linked data, nor to maintain the underlying server infrastructure. Therefore there is a need for helping end users to (1) take advantage of the information readily available in the LOD cloud for analyzing there own data and (2) contribute back to the linked data by representing their data and the associated metadata in the proper format as well as by linking them to other resources. In this context, the main originality is the graphical interface that allows any SPARQL query to be built transparently and iteratively by a non-expert user.

**Application** This software was developed in the context of the MirnAdapt (pea-aphid) project in 2016. The tool has been presented to the agriculture communities in conferences [53], [84] and to the Galaxy community [29]. Up to now, more than 10 biological partners team are actually testing and using the prototype software (colza, pea-aphids, copper microbiology, marine biology), and SANOFI has shown its interest to co-develop the tool. Even if its current user base belongs to the bioinformatics community, the scope of *AskOmics* is domain-independent and has the potential to reach a wider audience related to the Semantic Web community.

### 3.5.2. AuReMe workspace

Goal Tracable reconstruction of metabolic networks [url].

**Description** The toolbox *AuReMe* allows for the Automatic Reconstruction of Metabolic networks based on the combination of multiple heterogeneous data and knowledge sources [64]. It is available as a Docker image. Five modules are composing *AuReMe*: 1) The Model-management PADmet module allows manipulating and traceing all metabolic data via a local database. [package] 2) The meneco python package allows the gaps of a metabolic network to be filled by using a topological approach that implements a logical programming approach to solve a combinatorial problem [107], [65] and [21] [python package] 3) The shogen python package allows genome and metabolic network to be aligned in order to identify genome units which contain a large density of genes coding for enzymes; it also implements a logical programming approach [60] [python package]. 4) The manual curation assistance PADmet module allows the reported metabolic networks and their metadata to be curated. 5) The Wiki-export PADmet module enables the export of the metabolic network and its functional genomic unit as a local wiki platform allowing a user-friendly investigation [package].

**Originality** The main added-values are the inclusion of graph-based gap-filling tools that are particularly relevant for the study of non-classical organisms, the possibility to trace the reconstruction and curation procedures, and the representation and exploration of reconstructed metabolic networks with wikis.

**Application** The tools included in *AuReMe* have been used for reconstructing metabolic networks of micro and macro-algae [97], extremophile bacteria [16] and communities of organisms [61] in the context of the Idealg, Ciric-omics and IPL Algae-In-Silico projects.

### 3.5.3. FinGoc-tools

Goal Filtering interaction networks with graph-based optimization criteria.

**Description** The goal is to offer a set of tools for the reconstruction of networks from genome, literature and large-scale observation data (expression data, metabolomics...) in order to elucidate the main regulators of an observed phenotype. Most of the optimization issues are addressed with Answer Set Programming. 1) The lombarde package enables the filtering of transcription-factor/binding-site regulatory networks with mutual information reported by the response to environmental perturbations. The high level of falsepositive interactions is filters according to graph-based criteria. Knowledge about regulatory modules such as operons or the output of the shogen package can be taken into account [39], [38] [web server]. 2) The KeyRegulatorFinder package allows searching key regulators of lists of molecules (like metabolites, enzymes or genes) by taking advantage of knowledge databases in cell metabolism and signaling. The complete information is transcribed into a large-scale interaction graph which is filtered to report the most significant upstream regulators of the considered list of molecules [59] [package]. 3) The powerGrasp python package provides an implementation of graph compression methods oriented toward visualization, and based on power graph analysis. [package]. 4) The iggy package enables the repairing of an interaction graph with respect to expression data. It proposes a range of different operations for altering experimental data and/or a biological network in order to re-establish their mutual consistency, an indispensable prerequisite for automated prediction. For accomplishing repair and prediction, we take advantage of the distinguished modeling and reasoning capacities of Answer Set Programming. [6] [114] [Python package]

**Originality** The main added-value of these tools is to make explicit the criteria used to highlight the role of the main regulators: the underlying methods encode explicit graph-based criteria instead of relying on statistical approaches. This makes it possible to explain local relationships and patterns within interaction graphs by explicit biological relationships.

**Application** The tools have been used to figure out the main gene-regulators of the response of porks to several diets in [74], [76] and [18]. The tools were also used to to decipher regulators of reproduction for the pea aphid, an insect that is a pest on plants [85], [119].

### 3.5.4. Caspo software

Participant: Anne Siegel.

Goal Studying synchronous boolean networks [url]

**Description** Cell ASP Optimizer (*Caspo*) constitutes a pipeline for automated reasoning on logical signaling networks. The main underlying issue is that inherent experimental noise is considered, many different logical networks can be compatible with a set of experimental observations (see [106] and [22]). It is available as a Docker container. Five modules are composing *Caspo*: 1) the *Caspo-learn* module performs an automated inference of logical networks from experimental data allows for identifying admissible large-scale logic models saving a lot of efforts and without any a priori bias [115] and [78]. 2) The *Caspo-classify*, predict and visualize modules allows for classifying a family of boolean networks with respect to their input-output predictions [78]. 3) The *Caspo-design* module designs experimental perturbations which would allow for an optimal discrimination of rival models in a family of boolean networks [116]. 4) The *Caspo-control* module identifies key-players of a family of networks: it computes robust intervention strategies that force a set of target species or compounds into a desired steady state [80]. 5) The *Caspo-timeseries* module to take into account time-series observation datasets in the learning procedure [94] [python package and docker container].

**Originality** The *Caspo* modules provide friendly and efficient solutions to problems that were previously addressed in theoretical papers with MILP programs. The main advantage is that is enables a complete study of logical network without requiring any linear constraint programs.

**Application** The *Caspo* tool was initiated in the framework of the BioTempo project. *Caspo-learn* has been included as a module to learn logical networks from early steady-state data in CellNopt, a generic platform which implements several methods for learning and studying signaling networks are different modeling levels (from logical models to numerical models).

### 3.5.5. Cadbiom package

Goal Building and analyzing the asynchronous dynamics of enriched logical networks [url]

**Description** Based on Guarded transition semantic, the *Cadbiom* software provides a formal framework to help the modeling of biological systems such as cell signaling network. It allows synchronization events to be investigated in biological networks [40]. It is available as a Docker image. Three modules are composing *Cadbiom*: 1) The *Cadbiom* graphical interface is useful to build and study moderate size models. It provides exploration, simulation and checking. For large-scale models, *Cadbiom* also allows to focus on specific nodes of interest. 2) The *Cadbiom* API allows a model to be loaded, performing static analysis and checking temporal properties on a finite horizon in the future or in the past. 3) Exploring large-scale knowledge repositories, the translations of the large-scale PID repository (about 10,000 curated interactions) have been translated into the *Cadbiom* formalism.

**Originality** Model-checking approaches applied to Boolean networks [81] or multivalued networks [91] allow the trajectories of the system to be entirely studied but they can only be applied to small-size networks. On the contrary, *Cadbiom* is able to handle large-scale knowledge databases.

**Application** The *Cadbiom* tool was applied to study the regulators of the TGF- $\beta$ , a gene that controls liver fibrosis [40] in the framework of the *TGFSysBio* project. The study of its predictions also enabled large-scale knowledge databases (PID) to be curated [25].

### 3.5.6. Logol software

Goal Complex pattern modelling and matching [url]

**Description** The *Logol* toolbox is a swiss-army-knife for pattern matching on DNA/RNA/Protein sequences, using a high-level grammatical formalism to permit a large expressivity for patterns [50]. A *Logol* pattern can consist in a complex combination of motifs (such as degenerated strings) and structures (such as imperfect stem-loop ou repeats). *Logol* key features are the possibilities to divide a pattern description into several sub-patterns, to model long range dependencies, to enable the use of ambiguous models or to permit the inclusion of negative conditions in a pattern definition. The *LogolMatch* parser takes as input a biological sequence and a grammar file. It returns a XML file containing all the occurrences of the pattern in the sequence with their parsing details. The input sequences can be genomes from biological banks.

**Originality** Many pattern matching tools exist to efficiently model specific types of patterns: vmatch [82], patmatch [121], cutadapt [89], scoring matrix or profile HMMs [44], [71]. The main advantage of *Logol* is its very large expressivity. It encompasses most of the features of these specialized tools and enables interplays between several classes of patterns (motifs and structures).

**Application** The *Logol* tool was applied to the detection of mutated primers in a metabarcoding study [41], [42] or to stem-loop identification (e.g. in CRISPR <sup>0</sup> [103], [50]). Ongoing application is the search for transposable elements in the human genome in the context of a colorectal cancer study <sup>0</sup>. *Logol* strongly supported the study of the LXR- $\alpha$  targets in the framework of the FatInteger project.

### 3.5.7. Protomata-suite

Goal Expressive pattern discovery on protein sequences [url]

<sup>&</sup>lt;sup>0</sup>http://crispi.genouest.org/, https://hal.inria.fr/hal-00643408

<sup>&</sup>lt;sup>0</sup>Preprint: http://www.biorxiv.org/content/early/2017/03/09/115030

**Description***Protomata* is a machine learning suite for the inference of automata characterizing (functional) families of proteins from available sequences. Based on partial and local alignments, *Protomata* learns precise characterizations of the families of proteins, allowing new family members to be predicted with a high specificity. Three main modules are integrated in the *Protomata-learner* workflow are available as well as stand-alone programs: 1) *paloma* builds partial local multiple alignments, 2) *protobuild* infers automata from these alignements and 3) *protomatch* and *protoalign* scans, parses and aligns new sequences with learnt automata. The suite is completed by tools to handle or visualize data and can be used online by the biologists via a web interface on Genouest Platform. It is actively maintained (version v2.1 was released in April 2017) and we are scheduling a new major version with enhanced scoring schemes that we have proposed [105].

**Originality** The main specificity is that the power of characterization is beyond the scope of classical sequence patterns such as PSSM (e.g. MEME suite [43]), Profile HMM (e.g. HMMER package [71]), or Prosite Patterns [111] allowing new family members to be predicted with a high specificity.

**Application** The *Protomata* tool is used both to update automatically the Cyanolase database [62] and, when combined to Formal Concept Analysis, to automated enzyme classification, such as the HAD superfamily of proteins [67] in the framework of the *Idealg project*.

## 4. Application Domains

### 4.1. Application fields in biology

Our methods are applied in several fields of molecular biology.

Our main application field is **marine biology**, as it is a transversal field with respect to issues in integrative biology, dynamical systems and sequence analysis. Our main collaborators work at the Station Biologique de Roscoff. We are strongly involved in the study of brown algae: the *meneco, memap and memerge* tools were designed to realize a complete reconstruction of metabolic networks for non-benchmark species [97], [69]. On the same application model, the pattern discovery tool *protomata learner* combined with supervised bi-clustering based on formal concept analysis allows for the classification of sub-families of specific proteins [67]. The same tool also allowed us to gain a better understanding of cyanobacteria proteins [4]. At the larger level of 4D structures, classification technics have also allowed us to introduce new methods for the characterization of viruses in marine metagenomic sample [72]. Finally, in dynamical systems, we use asymptotic analysis (tool *pogg*) to decipher the initiation of sea urchin translation [51] [95]. We are currently involved in two new applications in this domain: the team participates to a Inria Project Lab program with the Biocore and Ange Inria teams, focused on the understanding on green micro-algae; and we are involved in the deciphering of phytoplancton variability at the system biology level in collaboration with the Station Biologique de Roscoff (ANR Samosa).

In **micro-biology**, our main issue is the understanding of bacteria living in extreme environments, mainly in collaboration with the group of bioinformatics at Universidad de Chile (funded by CMM, CRG and Inria-Chile). In order to elucidate the main characteristics of these bacteria, we develop efficient methods to identify the main groups of regulators for their specific response in their living environment. To that purpose, we use constraints-based modeling and combinatorial optimization. The integrative biology tools *meneco bioquali, ingranalysis, shogen, lombarde* were designed in this context [77]. In 2016, two applications focused on the study of extremophile consortium of bacteria have been performed with these tools [61], [38]. In parallel, in collaboration with Ifremer (Brest), we have conducted similar work to decipher protein-protein interactions within archebacteria [96]. Our sequence analysis tool (*logol*) allowed us to build and maintain a very expressive CRISPR database [10] [50].

Similarly, in **environmental sciences**, our goal is to propose methods to identify regulators of very complex phenotypes related to environmental issues. In collaboration with researchers from Inra/Pegase laboratory, we develop methods to distinguish the response of breeding animals to different diaries or treatments [37] and characterize upstream transcriptional regulators [59], applied to porks [74], [75], [76]. Semantic-based analysis was useful for interpreting differences of gene expression in pork meat [79].

In addition, constraints-based programming also allows us to decipher regulators of reproduction for the pea aphid, an insect that is a pest on plants [85], [120]. This was performed in collaboration with Inra/Igepp. This paved the way to the recent research track initiated in the team about integration of heterogeneous data with RDF-technologies (see AskOmics software) [84], [70] and about graph-compression (see powergrasp software).

In **bio-medical applications**, we focus our attention on the confrontation of large-scale measurements with large-scale knowledge repositories about regulation pathways such as Transpath, PID or pathway commons. In collaboration with Institut Curie, we have studied the Ewing Sarcoma regulation network to test the capability of our tool *bioquali* to accurately correct and predict a large-scale network behavior [47]. Our ongoing studies in this field focus on the exhaustive learning of discrete dynamical networks matching with experimental data, as a case study for modeling experimental design with constraints-based approaches. To that purpose, we collaborate with J. Saez Rodriguez group at EBI [115] and N. Theret group at Inserm/Irset (Rennes) [40]. The dynamical system tools *caspo and cadbiom* were designed within these collaborations. Ongoing studies focus on the understanding of the metabolism of xenobiotics (mecagenotox program) and the filtering of sets of regulatory compounds within large-scale signaling network (TGFSysBio project).

# 5. Highlights of the Year

### 5.1. Highlights of the Year

### 5.1.1. Awards

The team received a best paper award at the conference ICFCA and a best student paper award at the conference LPNMR.

**BEST PAPERS AWARDS :** 

[24] International Conference on Formal Concept Analysis 2017. L. BOURNEUF, J. NICOLAS.

[26] 14th International Conference on Logic Programming and Nonmonotonic Reasoning - LPNMR 2017. C. FRIOUX, T. SCHAUB, S. SCHELLHORN, A. SIEGEL, P. WANKO.

# 6. New Software and Platforms

### 6.1. AskOmics

KEYWORDS: RDF - SPARQL - Querying - Graph - LOD - Linked open data

FUNCTIONAL DESCRIPTION: AskOmics allows to load heterogeneous bioinformatics data (formatted as tabular files) into a Triple Store system using a user-friendly web interface. AskOmics also provides an intuitive graph-based user interface supporting the creation of complex queries that currently require hours of manual searches across tens of spreadsheet files. The elements of interest selected in the graph are then automatically converted into a SPARQL query that is executed on the users' data.

NEWS OF THE YEAR: Several functionalities have been developed: 1) capacity of integrating genomics data (import of GFF and BED files and generation of RDF compliant with the FALDO ontology), 2) integration of data and knowledge in the OWL format to exploit biological information from external repositories, particularly from EBI and NCBI. Notably, this functionality allows AskOmics to support the Gene Ontology, the Taxonomy ontology as well as BioPAX biological networks. 3) improved user interface expressivity for generating SPARQL queries, 4) implementation of a support for multiple concurrent user sessions, with the distcintion between public and user-specific datasets 5) deployment of AskOmics on the GenOuest cloud infrastructure to facilitate its release and diffusion 6) interoperability between AskOmics and the Galaxy workflow environment.

- Authors: Charles Bettembourg, Xavier Garnier, Anthony Bretaudeau, Fabrice Legeai, Olivier Dameron, Olivier Filangi and Yvanne Chaussin
- Partners: Université de Rennes 1 CNRS INRA
- Contact: Fabrice Legeai
- URL: https://github.com/askomics/askomics

### 6.2. PADMet-utils

KEYWORDS: Metabolic networks - Bioinformatics - Workflow - Omic data - Toolbox - Data management - LOD - Linked open data

FUNCTIONAL DESCRIPTION: The main concept underlying padmet-utils is to provide solutions that ensure the consistency, the internal standardization and the reconciliation of the information used within any workflow that combines several tools involving metabolic networks reconstruction or analysis.

NEWS OF THE YEAR: In 2017, Padmet-utils was enriched with a RDF export to allow the interoperability of the AuReMe workspace for the reconstruction of metabolic networks with the Askomics Tool for querying heterogeneous data. Padmet-utils was also extended to handle metabolic networks in the SBML3 format.

- Participants: Alejandro Maass, Meziane Aite and Anne Siegel
- Partner: University of Chile
- Contact: Anne Siegel
- URL: https://gitlab.inria.fr/maite/padmet-utils

### 6.3. CADBIOM

### Computer Aided Design of Biological Models

KEYWORDS: Health - Biology - Biotechnology - Bioinformatics - Systems Biology

FUNCTIONAL DESCRIPTION: Based on Guarded transition semantic, this software provides a formal framework to help the modeling of biological systems such as cell signaling network. It allows investigating synchronization events in biological networks. Software development has been restarted since November 2016. The source code is available at the following address: https://gitlab.irisa.fr/0000B8EG/Cadbiom

- Participants: Geoffroy Andrieux, Michel Le Borgne, Nathalie Theret, Nolwenn Le Meur and Pierre Vignet
- Contact: Anne Siegel
- URL: http://cadbiom.genouest.org

### 6.4. conquests

### Crossroads in Metabolic Network from Stoechiometric and Topologic Studies

KEYWORDS: Bioinformatics - ASP - Answer Set Programming - Constraint-based programming FUNCTIONAL DESCRIPTION: This Python package in systems biology allows the identification of essential metabolites with respect to the production of targeted elements in a metabolic network, by comparing flux and graph-based analysis. Conquests's inputs are a sbml file corresponding to a metabolic network and the biomass reaction name. The outputs are three sets of essential metabolites. They are computed according to three complementary criteria: graph-based accessibility of targeted metabolites, the presence of flux in the biomass reaction and the maximisation of flux in the biomass reaction.

NEWS OF THE YEAR: Conquest was released in 2017.

• Contact: Julie Laniau

# 7. New Results

### 7.1. Data integration and pre-processing with semantic-based technologies

Participants: Olivier Dameron, Xavier Garnier, Yann Rivault, Anne Siegel, Denis Tagu.

Interoperable infrastructure and implementation of a health data model for remote monitoring of chronic diseases with comorbidities In the context of telemedecine, we worked on a numerical application for monitoring patients with chronic diseases. We have developed a system based on a formal ontology that integrates the alert information and the patient data extracted from the electronic health record in order to better classify the importance of alerts. A pilot study was conducted on atrial fibrillation alerts. The results suggest that this approach has the potential to significantly reduce the alert burden in telecardiology [101], [100]. In 2017, we proposed an architecture supporting data exchange in the context of multiple chronic diseases [*O. Dameron, Y. Rivault*] [27].

AskOmics, a web tool to integrate and query biological data using semantic web technologies The software AksOmics has been adapted to two types of scientific topics important in agronomical and environmental sciences: plant genomic data and insect pest genomic data. With *AskOmics*, plant genomicists (from academic and private labs from the Rapsodyn project - Investment for the future) working on the rapeseed (Brassica napus) are able to tackle the understanding of which gene copy is active or repressed in key developmental processes in relation with seed quality and oil production, in the frame of plant breeding. Additionally, entomologists use this tool to extract valuable knowledge on the way insect pests such as aphids are able to rapidly disseminate on crops, in the frame of free-pesticide methods for plant protection. *AskOmics* has been presented to the international community of insect genomics (i5k: http://i5k.github.io/ ) by web-seminars and *AskOmics* developers have been invited at international workshops. For facilitating AskOmics's adoption by end-user, it has recently been integrated within the Galaxy workflow engine [*O. Dameron, X. Garnier, A. Siegel, D. Tagu*] [28], [29], [23]

### 7.2. Data and knowledge integration based on combinatorial optimization

**Participants:** Meziane Aite, Lucas Bourneuf, Marie Chevallier, Damien Eveillard, Clémence Frioux, Jeanne Got, Julie Laniau, François Moreews, Jacques Nicolas, Anne Siegel.

A transcriptome multi-tissue analysis identifies biological pathways and genes associated with variations in feed efficiency of growing pigs Our work on the identification of upstream regulators within large-scale knowledge databases (prototype *KeyRegulatorFinder*) [59] was valuable for figuring out the main generegulators of the response of porks to several diets [*F. Moreews, A. Siegel*] [18]

FCA in a Logical Programming Setting for Visualization-oriented Graph Compression We have explored the underlying idea of lossless network compression to address the problem of uncertainty in biological networks built from predictions, to help to visualize the networks and to classify their nodes in accordance with available annotations [119]. Network compression has been used with success in Dresden (M. Schroeder) with a heuristic approach called Power Graph analysis building abstract graphs where nodes are clusters of nodes in the initial graph and edges represent bicliques between two sets of nodes. First encouraging results have been presented (best paper award) showing that it is possible to mimic the Power Graph behaviour while opening the possibility to achieve better compression levels compared to alternative compression schema. [*L. Bourneuf, J. Nicolas*] [24]

**Metabolic network completion and analysis** We released the application paper of the tool *Meneco*, a tool dedicated to the topological gap-filling of genome-scale draft metabolic networks. The tool reformulates gap-filling as a qualitative combinatorial optimization problem, omitting constraints raised by the stoichiometry, and solves this problem using Answer Set Programming. Run on an artificial test set of 10,800 degraded *Escherichia coli* networks, we evidenced that *Meneco* outperforms the stoichiometry-based tool *Gapfill* in terms of precision. In addition, *Meneco* reports 10 times less putative reactions than MILP-based tool *Fastgapfill* for an equivalent precision. This is a strong advantage for manual curation post-processing, since curating 50 to 80 reactions is still possible whereas manually-curating 800 reactions is out-of-range. Meneco was applied to the reconstruction and understanding of a pathogeneic strain of salmon. [*C. Frioux, J. Got, A. Siegel*] [21], [16]

**Toward the study of metabolic functions in communities of organisms** In [21], we provided a first example on how to use topological metabolic modeling to assess the complementarity between two members of an algal ecosystem. Since this study, we generalized the selection of subcommunities of interest and propose likely interactions that could occur between seaweeds and their associated bacteria. A focus has also been done on plant microbiota and the reasons underlying the organization of the community. Altogether, these on-going works enable a better understanding of holobiont organizations and functioning. [*M. Aite, M. Chevallier, C. Frioux, J. got, A. Siegel, C. Trottier*] [21], [31], [30]

**Hybrid Metabolic Network Completion** In order to improve the precision of gap-filling approaches, we introduced a hybrid approach to formally reconcile existing stoichiometric and topological approaches to network completion in a unified formalism. An hybrid ASP encoding based on MILP constraint propagator

was developed. It relies upon the theory reasoning capacities of the ASP system Clingo to solve the resulting logic program with linear constraints over reals. For short, this technology made it possible to combine the best of the combinatorial problem solver Clingo with the MILP solver CPlex. Run on the artificial test set of 10,800 degraded *Escherichia coli* networks introduced in [21], our approach yielded greatly superior results than obtainable from purely qualitative or MILP approaches. [*C. Frioux, A. Siegel*] [26], [19]

**Combining graph and flux-based structures to decipher phenotypic essential metabolites within metabolic networks** Whenever flux or graph-based criteria are used to study metabolic networks, these analyses are generally centered on the outcome of the network and considers all metabolic compounds to be equivalent in this respect. We generalized the concept of essentiality to metabolites and introduced the concept of the phenotypic essential metabolite (PEM) which influences the growth phenotype according to sustainability, producibility or optimal-efficiency criteria. The exhaustive study of phenotypic essential metabolites in six genome-scale metabolic models suggests that the combination and the comparison of graph, stoichiometry and optimal flux-based criteria allow some features of the metabolic network functionality to be deciphered by focusing on a small number of compounds. [*C. Frioux, J. Laniau, A. Siegel*] [19]

### 7.3. Systems biology

**Participants:** Jérémie Bourdon, Jean Coquet, Victorien Delannée, Jacques Nicolas, Anne Siegel, Nathalie Théret, Pierre Vignet.

A modeling approach to evaluate the balance between bioactivation and detoxification of MeIQx in human hepatocytes Heterocyclic aromatic amines (HAA), including MeIQx, are environmental and food contaminants that are potentially carcinogenic for humans. Using a computational approach, we developed a numerical model for MeIQx metabolism that predicts the MeIQx biotransformation into detoxification or bioactivation pathways according to the concentration of MeIQx. Our results demonstrate that CYP1A2 is a key enzyme in the system that regulates the balance between bioactivation and detoxification. This highlights the importance of complex regulations of enzyme competitions that should be taken into account in any multiorgan model [*V. Delannée, A. Siegel, N. Théret*] [17]

**caspo:** a toolbox for automated reasoning on the response of logical signaling networks families The accompanying paper of the complete family of modules introduced in the caspo software was published in 2017 (see software section for details) [*A. Siegel*] [22]

Identifying Functional Families of Trajectories in Biological Pathways by Soft Clustering: Application to TGF- $\beta$  Signaling At a dynamical level, in [40], reaction-based and regulatory information was transpoed in a unified formalism of enriched Petri Nets (discrete dynamical systems), namely a simplified version of guarded transitions in which we introduced temporal parameters for each transition to manage competition and cooperation between parts of the models. This allowed integrating the 137 human signaling maps from the Pathway Interaction Database (PID) into a single unified large-scale dynamic model. Simulation and model checking analyses evidence that 15,934 different sets of molecules are able to regulate 159 of TGF- $\beta$  target genes (TGF- $\beta$  is a multifunctional cytokine that regulates mammalian cell development, differentiation, and homeostasis). Further analysis of these 15,934 sets of molecules by biological experts is obviously impractical. Our study identified five clusters of sets of molecules for which enrichment analysis highlighted the overrepresented molecules as well as the specific biological processes they are associated with. These results are biologically-relevant and consistent with the pleiotropic nature of TGF- $\beta$  [*J. Coquet, N. Théret, O. Dameron*] [25]

A Logic for checking the probabilistic steady-state properties of reaction networks. We have constructed a probabilistic analog to flux balance analysis of reaction networks to enable a formal verification of logical constrains about the stationary regime of a system by using information from experimental variances and co-variances. This is mainly based on a stationary analysis of the probabilistic dynamics relying on a Bernoulli approximation of a reaction network. The analysis requires solving non linear optimization problems [*J. Bourdon, A. Siegel*] [20]

### 7.4. Sequence and structure annotation

Participants: Catherine Belleannée, François Coste, Jacques Nicolas.

Better scoring schemes for the recognition of functional proteins by protomata The machine learning algorithm included in *Protomata-learner* learns weighted automata representing both functional families from the sequences of amino acids, and the possible disjunctions between members. We investigated alternative sequence weighting strategies and null-models. We introduced a normalization of the score, and a method to assess the significance of scores, to simplify the prediction. Preliminary results show a good improvement of the prediction power of the computed models. [*F. Coste*] [36]

**Detection of mutated primers and impact on targeted metagenomics results** In targeted metagenomics, an initial task is the detection in each sequence of the primers used for amplifying the targeted region. The selected sequences are then trimmed and clustered in order to inventory the species present in the sample. Common pratices consist in retaining only the sequences with perfect primers (i.e. non-mutated by sequencing error). In the context of a study characterizing the biodiversity of tropical soils in unicellular eukaryotes, we have implemented the search for mutated primers, using the grammatical pattern matching tool Logol, and shown that retrieving sequences with mutated primers has a significant impact on targeted metagenomics results, as it makes possible to detect more species (7% additional OTUs in our study). [*C. Belleannée*] [34].

**First landscape of binding to chromosomes for a domesticated mariner transposase in the human genome.** In order to study the diversity of genomic targets of the SETMAR protein in two colorectal cell lines, a first task was to massively discover the Made1 80-bp transposon element in the human genome. For that, we used our Logol grammar-like approach to look for non perfect Made1 instances. In Logol, a pattern can be divided into several sub-patterns. The Made1 model took advantage of this feature to strengthen the most conserved regions. Cumulating this search with the Blast alignment search permitted to significantly increase the Made1 annotation in the human genome.[*C. Belleannée*] [33]

### 8. Bilateral Contracts and Grants with Industry

### 8.1. Bilateral Contracts with Industry

Participants: Olivier Dameron, Anne Siegel, Méline Wery.

Our software *AskOmics* was considered as relevant by the SANOFI bio-medical company in order to facilitate the integration and the query of the data produced by their scientists. A former Ph.D. of Dyliss who designed the first prototypes of *AskOmics* was recruited by SANOFI. Since then, SANOFI is included in the developer's team of *AskOmics* and a joint Dyliss–SANOFI CIFRE Ph.D. thesis started about the integration of complementary reasoning features to SPARQL queries in Oct. 2017.

### 9. Partnerships and Cooperations

### 9.1. Regional Initiatives

### 9.1.1. Regional initiative: the Ecosyst project

Participants: Damien Eveillard, Marie Chevallier, Clémence Frioux, Anne Siegel, Camille Trottier.

EcoSyst is a Biogenouest inter-regional federating project (Brittany & Pays de la Loire) aiming at the emergence of Systems Ecology at the level of western France regions. Drawing on the strengths and skills involved, EcoSyst targets the incubation of new ideas and new projects at disciplinary interfaces. Thanks to this community project, we want to develop the skills of Ecology, Environment, Modeling, Bioinformatics and Systems Biology and their application to organisms and ecosystems of interest in agronomy, sea and health. EcoSyst includes also the identification of the major issues and concerns, the fundamental and essential methods and the very real needs of the community (training, tools, ...); this in order to consider the construction of a community platform (or an offer of service within an existing platform) on complex systems modeling, meeting expectations of the community as fully as possible.

### 9.1.2. Regional partnership with computer science laboratories in Nantes

**Participants:** Anne Siegel, Jérémie Bourdon, Damien Eveillard, François Coste, Maxime Folschette, Jacques Nicolas.

Methodologies are developed in close collaboration with the LS2N (fusion of LINA and IRCCyN) located at University of Nantes and École centrale de Nantes. This is acted through the Biotempo and Idealg ANR projects and co-development of common software toolboxes within the Renabi-GO platform support. C. Trottier is a co-supervised bioianalysis and software development engineer within the Idealg project. M. Chevallier is a co-supervised development and animation engineer within the regional initiative "Ecosyst". In addition, the ongoing Ph-D student J. Laniau is co-supervised with a member of the LS2N laboratory. Finally, M. Folschette is a PostDoc working on on a project aiming at analyzing TGF-beta-related pathways evolutions after epithelial-mesenchymal transition in liver cancer, which is a recognized biological process leading to metastasis. This project is based on a topic shared with the LS2N: the use of graph coloring and reconstruction to witness expression changes, and is funded by the Université Bretagne Loire.

### 9.1.3. Regional partnership in Marine Biology

**Participants:** Meziane Aite, Arnaud Belcour, Catherine Belleannée, Jérémie Bourdon, Jean Coquet, François Coste, Damien Eveillard, Olivier Dameron, Clémence Frioux, Jeanne Got, Julie Laniau, Jacques Nicolas, Camille Trottier, Anne Siegel.

A strong application domain of the Dyliss project is marine Biology. This application domain is co-developed with the station biologique de Roscoff and their three UMR and involves several contracts. Our approach based on parcimonious modelling allowed an in silico characterization of processes required within sea urchin translation [83], [95]. We are also strongly involved in the the IDEALG consortium, a long term project (10 years, ANR Investissement avenir) aiming at the development of macro-algae biotechnology. Among the research activities, we are particularly interested in the analysis and reconstruction of metabolism and the characterization of key enzymes. Our methods based on combinatorial optimization for the reconstruction of genome-scale metabolic networks and on classification of enzyme families based on local and partial alignments allowed the *E. Siliculosus*eaweed metabolism to be deciphered [97], [67]. As a further study, we reconstructed the metabolic network of a symbiot bacterium *Ca. P. ectocarpi* [69] and used this reconstructed network to decipher interactions within the algal-bacteria holobiont [21].

### 9.1.4. Regional partnership in agriculture and environmental sciences

**Participants:** Catherine Belleannée, François Coste, Olivier Dameron, Xavier Garnier, François Moreews, Jacques Nicolas, Anne Siegel, Denis Tagu.

We have a strong and long term collaboration with biologists of INRA in Rennes : PEGASE and IGEPP units. F. Morrews is a permanent engineer from PEGASE center hosted in the team to develop methods for integrative biology applied to species of interest in agriculture. D. Tagu is a research director at INRA/IGEPP who spends 20% of his time in the team to develop collaborative projects. This partnership has been supported by the co-supervision of phDs, post-docs and engineers. This collaboration was also reinforced by collaboration within ANR contracts (MirNadapt, FatInteger).

In collaboration with researchers from the PEGASE center (INRA) focused on breeding animals, we have contributed to several studies aiming at better integrating and investigating data in order to facilitate animal selection and alimentation. The *NutritionAnalyzer* prototype was developed to understand better the impact of several diaries or treatments for lactary cows over the composition of milk [37]. Our work on the identification of upstream regulators within large-scale knowledge databases (prototype *KeyRegulatorFinder*) [59] and on semantic-based analysis of metabolic networks [54] was also very valuable for interpreting differences of gene expression in pork meat [79] and figure out the main gene-regulators of the response of porks to several diets (see [74], [76] and [18]).

In addition, constraints-based programming also allows us to decipher regulators of reproduction for the pea aphid, an insect that is a pest on plants in the framework of the *MirnAdapt* project. In terms of biological output of the network studies on the pea aphid microRNAs, we have identified one new microRNA (apmir-3019, not present in any known species other than the pea aphid) who has more than 900 putative mRNA targets. All these targets, as well as apmir3019, are differentially expressed between sexual and asexual embryos [85], [119].

### 9.1.5. Regional partnership in health

**Participants:** Jean Coquet, Olivier Dameron, Victorien Delannée, Marine Louarn, Anne Siegel, Nathalie Théret, Pierre Vignet.

We also have a strong and long term collaboration in health, namely with the IRSET laboratory at Univ. Rennes 1. N. Théret, research director at INSERM, is hosted in the team to strenghen our collaborative projects. Our collaborations are acted by the co-supervised Ph-D theses of V. Delannée [14], M. Conan (Metagenotox project, funded by Anses) and J. Coquet [12]. This partnership was reinforced by the ANR contract Biotempo ended at the end of 2014. In 2015, the project of combining semantic web technologies and bi-clustering classification based on formal concept analysis was applied to systems biology within the PEPS CONFOCAL project. This scientific project has been recently pushed forward in the recent TGFSYSBio project funded by Plan Cancer on the modelling of the microenvironment of TGFbeta signaling network (P. Vignet has been recruited on this contract at the end of 2016).

A new application was initiated in 2017 through a collaboration with Rennes hospital, supported by a Inria-INSERM Ph-D thesis (M. Louarn).

### 9.2. National Initiatives

### 9.2.1. ANR Idealg

**Participants:** Meziane Aite, Arnaud Belcour, Jérémie Bourdon, Marie Chevallier, François Coste, Damien Eveillard, Clémence Frioux, Jeanne Got, Julie Laniau, Jacques Nicolas, Anne Siegel.

IDEALG is one of the five laureates from the national call 2010 for Biotechnology and Bioresource and will run until 2020. It gathers 18 different partners from the academic field (CNRS, IFREMER, UEB, UBO, UBS, ENSCR, University of Nantes, INRA, AgroCampus), the industrial field (C-WEED, Bezhin Rosko, Aleor, France Haliotis, DuPont) as well as a technical center specialized in seaweeds (CEVA) in order to foster biotechnology applications within the seaweed field. We are participating to the tasks related to the establishment of a virtual platform for integrating omics studies on seaweeds and the integrative analysis of seaweed metabolism, in cooperation with SBR Roscoff. Major objectives are the building of brown algae metabolic maps, flux analysis and the selection of symbiotic bacteria to brown algae. We will also contribute to the prediction of specific enzymes (sulfatases) [More details].

### 9.2.2. Programs funded by research institutions

# 9.2.2.1. PEPS PEPS: a platform for supporting studies in pharmaco-epidemiology using medico-administrative databases

### Participants: Olivier Dameron, Yann Rivault.

As a partner of the PEPS platform, several teams at Inria Rennes develop generic methods supporting efficient and semantically-rich queries for pharmaco-epidemiology studies on medico-administrative databases. The leader is Thomas Guyet (Inria team Lacodam). We showed that Semantic Web technologies are technically suited for representing patients' data from medico-administrative databases as RDF and querying them using SPARQL. We also demonstrated that this approach is relevant as it supports the combination of patients' data with hierarchical knowledge in order to address the problem of reconciling precise patients data with more general query criteria [45], [99], [98]. This work is mostly conducted by Yann Rivault, whose PhD thesis is supervized by Olivier Dameron and Nolwenn LeMeur (Ecole des Hautes Etudes en Santé Publique).

#### 9.2.2.2. Cancer Plan: TGFSysBio

**Participants:** Jean Coquet, Olivier Dameron, Maxime Folschette, Vijay Ingalalli, Jacques Nicolas, Anne Siegel, Nathalie Théret, Pierre Vignet.

The TGFSYSBIO project aims to develop the first model of extracellular and intracellular TGF-beta system that might permit to analyze the behaviors of TGF-beta activity during the course of liver tumor progression and to identify new biomarkers and potential therapeutic targets. Based on collaboration with Jérôme Feret from ENS, Paris, we will combine a rule-based model (Kappa language) to describe extracellular TGF-beta activation and large-scale state-transition based (Cadbiom formalism) model for TGF-beta-dependent intracellular signaling pathways. The multi-scale integrated model will be enriched with a large-scale analysis of liver tissues using shotgun proteomics to characterize protein networks from tumor microenvironment whose remodeling is responsible for extracellular activation of TGF-beta. The trajectories and upstream regulators of the final model will be analyzed with symbolic model checking techniques and abstract interpretation combined with causality analysis. Candidates will be classified with semantic-based approaches and symbolic bi-clustering technics. The project is funded by the national program "Plan Cancer - Systems biology" from 2015 to 2018.

### 9.2.2.3. ANR Samosa

Participants: Mael Conan, Damien Eveillard, Jeanne Got, Anne Siegel.

Oceans are particularly affected by global change, which can cause e.g. increases in average sea temperature and in UV radiation fluxes onto ocean surface or a shrinkage of nutrient-rich areas. This raises the question of the capacity of marine photosynthetic microorganisms to cope with these environmental changes both at short term (physiological plasticity) and long term (e.g. gene alterations or acquisitions causing changes in fitness in a specific niche). *Synechococcus* cyanobacteria are among the most pertinent biological models to tackle this question, because of their ubiquity and wide abundance in the field, which allows them to be studied at all levels of organization from genes to the global ocean.

The SAMOSA project is funded by ANR from 2014 to 2018, coordinated by F. Gaczarek at the Station Biologique de Roscoff/UPMC/CNRS. The goal of the project is to develop a systems biology approach to characterize and model the main acclimation (i.e., physiological) and adaptation (i.e. evolutionary) mechanisms involved in the differential responses of *Synechococcus* clades/ecotypes to environmental fluctuations, with the goal to better predict their respective adaptability, and hence dynamics and distribution, in the context of global change. For this purpose, following intensive omics experimental protocol driven by our colleagues from – Station Biologique de Roscoff –, we aim at constructing a gene network model sufficiently flexible to allow the integration of transcriptomic and physiological data.

#### 9.2.2.4. ANSES Mecagenotox

Participants: Victorien Delannée, Mael Conan, Anne Siegel, Nathalie Théret.

The objective of Mecagenotox project is to characterize and model the human liver ability to bioactivate environmental contaminants during liver chronic diseases in order to assess individual susceptibility to xenobiotics. Indeed, liver pathologies which result in the development of fibrosis are associated with a severe dysfunction of liver functions that may lead to increased susceptibility against contaminants. In this project funded by ANSES and coordinated by S. Langouet at IRSET/inserm (Univ. Rennes 1), we will combine cell biology approaches, biochemistry, biophysics, analytical chemistry and bioinformatics to 1) understand how the tension forces induced by the development of liver fibrosis alter the susceptibility of hepatocytes to certain genotoxic chemicals (especially Heterocyclic Aromatic Amines) and 2) model the behavior of xenobiotic metabolism during the liver fibrosis. Our main goal is to identify "sensitive" biomolecules in the network and to understand more comprehensively bioactivation of environmental contaminants involved in the onset of hepatocellular carcinoma.

### 9.2.3. Programs funded by Inria

### 9.2.3.1. ADT Complex-biomarkers and ADT Proof of concept

Participants: Jeanne Got, Marie Chevallier, Meziane Aite, Anne Siegel.

These projects started in Oct. 2014 and aims at designing a working environment based on workflows to assist molecular biologists to integrate large-scale omics data on non-classical species. The main goal of the workflows will be to facilitate the identification of set of regulators involved in the response of a species when challenged by an environmental stress. Applications target extremophile biotechnologies (biomining) and marine biology (micro-algae).

### 9.2.3.2. IPL Algae in silico

Participants: Meziane Aite, Jeanne Got, Julie Laniau, Anne Siegel.

Microalgae are recognized for the extraordinary diversity of molecules they can contain: proteins, lipids (for biofuel or long chain polyunsaturated fatty acids for human health), vitamins, antioxidants, pigments. The project aims at predicting and optimizing the productivity of microalgae. It involves mainly the inria teams BIOCORE (PI), ANGE and DYLISS. Dyliss is in charge of the identification of physiological functions for microalgae based on their proteomes, which is undergone through the reconstruction of the metabolic network of the *T. lutea* microalgae.

### 9.2.3.3. IPL Neuromarkers

Participants: Olivier Dameron, Anne Siegel.

The project aims at identifying the main markers of pathologies through the production and the integration of imaging and bioinformatics data. It involves mainly the inria teams Aramis (PI) Dyliss, Genscale and Bonsai. Dyliss is in charge of facilitating the interoperability of imaging and bioinformatics data.

### 9.2.3.4. FederatedQueryScaler (Exploratory Research Action)

Participants: Olivier Dameron, Xavier Garnier, Vijay Ingalalli.

This project aims at developing automatic generation of abstractions for biological data and knowledge in order to scale federated queries in the context of semantic web technologies. It is a common project with the WIMMICS Inria team.

### 9.3. European Initiatives

### 9.3.1. Collaborations with Major European Organizations

Partner: Aachen university (Germany)

Title: Modeling the logical response of a signalling network with constraints-programming.

### 9.4. International Initiatives

### 9.4.1. Inria International Labs

### 9.4.1.1. Other IIL projects

We have a cooperation with Univ. of Chile (MATHomics, A. Maass) on methods for the identification of biomarkers and software for biochip design. It aims at combining automatic reasoning on biological sequences and networks with probabilistic approaches to manage, explore and integrate large sets of heterogeneous omics data into networks of interactions allowing to produce biomarkers, with a main application to biomining bacteria. The program is co-funded by Inria and CORFO-chile from 2012 to 2016. In this context, Integrative-BioChile was an Associate Team between Dyliss and the Laboratory of Bioinformatics and Mathematics of the Genome hosted at Univ. of Chile funded from 2011 to 2016. The collaboration is now supported by Chilean programs.

### 9.5. International Research Visitors

### 9.5.1. Visits of International Scientists

- Niger. University of Maradi [O. Abdou-Arbi]
- Poland. Politechnika Wroclawska [W. Dyrka]
- India. VIT University, Vellore [K. Lakshmanan]

### 9.5.2. Visits to International Labs

• Chile. University of Chile [A. Siegel, C. Frioux]

### 9.5.2.1. Research Stays Abroad

• Germany. University of Potsdam [L. Bourneuf, 3 months (nov 2017 - jan 2018)]

# **10.** Dissemination

### **10.1. Promoting Scientific Activities**

### 10.1.1. Scientific Events Selection

10.1.1.1. Member of the Conference Program Committees

- SWAT4HCLS (2017) Semantic Web and Tools for Health Care and Life Sciences (O. Dameron)
- BBCC (2017): Bioinformatica e Biologia Computazionale in Campania (O. Dameron)
- JOBIM (2017): French conference of Bioinformatics (A. Siegel)
- SIIM (2017) Symposium sur l'Ingénierie des Informations Médicales (O. Dameron)

### 10.1.1.2. Review

• ISMB/ECCB 2017.

### 10.1.2. Journal

10.1.2.1. Member of the Editorial Boards

- O. Dameron is an associate editor of the Journal of Biomedical Semantics
- J. Bourdon in an academic editor of PLoS One

### 10.1.2.2. Reviewer - Reviewing Activities

- Briefings in Bioinformatics, (O. Dameron)
- Journal of Biomedical Semantics (O. Dameron)
- Journal on Data Semantics (O. Dameron)
- Molecular Cancer (J. Nicolas),
- Plos One (J. Nicolas)

### 10.1.3. Invited Talks

10.1.3.1. French seminars

Paris, Hopital Lariboisière (Seminar, 2017) – SANOFI (Gentilly, 3 Invited seminars, 2017) – Clermont-Ferrand (Insect team, 2017) – Nantes (Université of Nantes, 2017).

### 10.1.3.2. Workshops and meetings

Conference on Boolean networks (Marseille, 2017) – Bioss Meeting on artificial intelligence (Gif, 2017)
# 10.1.4. Leadership within the Scientific Community

- Member of the steering committee of the International Conference on Grammatical Inference.
- The team was involved in the foundation of a national working group on the symbolic study of dynamical systems named bioss [web access]. The group gathers more than 170 scientists, from computer science to biology. Three meetings were organized this year. The group is supported by two French National Research Networks: bioinformatics (GDR BIM : bioinformatique moléculaire) and informatics-mathematics (GDR IM : Informatique Mathématique). It gathered twice in 2017: for a general meeting in Montpellier (Mar. 2017) and for a workshop focused on links between systems biology and artificial intelligence in Orsay (June 2017).

# 10.1.5. Scientific Expertise

### 10.1.5.1. International responsibilities

• Evaluation panel of the "Europe-USA Call Strengthening Transnational Research in Molecular Plant Sciences" launched by ERA-CAPS.

### 10.1.5.2. National responsibilities

- Institutional boards for the recruitment and evaluation of researchers. Inria National evaluation board (A. Siegel, nominated member). National Council of Universities, section 65 (O. Dameron, nominated member).
- Evaluation committees of French laboratories or doctoral schools. Bioinformatics groups of Institut Curie (Paris, presidency of the committee, A. Siegel) Doctoral school of Nice University (N. Théret).
- Presidency of the expert panel for the call *Systems biology applied to Cancer* of the National Cancer Plan 2017 (A. Siegel).
- **Recruitment committees**. Inria Senior Researchers (national committee, A. Siegel) Inria Junior Researchers (Nice, National Committee, A. Siegel)
- Scientific Advisory boards GDR BIM " Molecular Bioinformatics" (J. Nicolas).

## 10.1.5.3. National scientific boards

- Scientific Advisory Board of the French National Research Network GDR BIM *Molecular Bioinformatics* (J. Nicolas).
- Operational Legal and Ethical Risk Assessment Committee (COERLE) at Inria (J. Nicolas).
- Animation of the Bioss working group (A. Siegel).
- Board of directors of the French Society for biology of the extracellular matrix (N. Théret).

## 10.1.5.4. Prospective working groups

- "Big & Open Data" foresight working group of PROSPER network (F. Coste).
- "Prospectives in predictive toxicology" working group at INRA (A. Siegel)

## 10.1.5.5. Local responsibilities

- Scientific Advisory Board of Biogenouest (J. Bourdon, N. Théret)
- IRISA laboratory (computer science department of Univ. Rennes 1) council (A. Siegel)
- Responsability of the IRISA laboratory "Health-biology" cross-cutting axis (O. Dameron)<sup>0</sup>
- SCAS (Service Commun d'Action Sociale) of Univ. Rennes 1 (C. Belleannée)
- Scientific committee of Univ. Rennes 1 school of medicine (O. Dameron, A. Siegel).

<sup>0</sup>http://www.irisa.fr/en/page/crosscutting-axis

# 10.2. Teaching - Supervision - Juries

### 10.2.1. Teaching track responsibilities

- Coordination of the doctoral school "Life, Agronomy and Health" of University of Rennes 1 [N. Théret]
- Coordination of the master degree "Bioinformatics and genomics", Univ. Rennes1 [O. Dameron]
- Coordination of the sub-domain "From Data to Knowledge: Machine Learning, Modeling and Indexing Multimedia Contents and Symbolic Data", Master in Computer Science, University of Rennes 1, France [F. Coste].

## 10.2.2. Course responsibilities

- "Atelier bioinformatique", Licence 2 informatique, Univ. Rennes 1 [O. Dameron]
- "Bioinformatique pour la génomique", 2nd year school of medicine, Univ. Rennes 1 [O. Dameron]
- "Bases de mathématiques et probablité" and "Méthodes en informatique", Master1 in public health, Univ. Rennes 1 [O. Dameron]
- "Big data and Semantic Web", Master 2 in public health, Univ. Rennes 1 [O. Dameron]
- "Intégration: Remise à niveau en informatique", Master 1 in bioinformatics, Univ. Rennes 1 [O. Dameron]
- "Programmation en Python", Master 1 in Public Health, Univ. Rennes 1 [O. Dameron]
- "Programmation impérative en Python", Master 1 in bioinformatics, Univ. Rennes 1 [O. Dameron]
- "Système informatique GNU/Linux", Master 1 in bioinformatics, Univ. Rennes 1 [O. Dameron]
- "Semantic Web and bio-ontologies", Master 2 in bioinformatics, Univ. Rennes 1 [O. Dameron]
- "e-Santé et réseaux hospitaliers", last year in engineering school ESIR, Univ. Rennes 1, [O. Dameron]
- "Equilibre Dynamique de la communication Cellulaire" Master 2 in Sciences cellulaire et Moléculaire du Vivant, Univ. Rennes 1 [N. Theret]

### 10.2.3. Teaching

Licence: C. Belleannée, Langages formels, 20h, L3 informatique, Univ. Rennes1, France. Licence: C. Belleannée, Algorithmique et Programmation Fonctionnelle, 60, L1 informatique, Univ. Rennes1, France.

Licence: J. Coquet, Module Programmation Scientifique 1, 20h, L1 informatique, Rennes1, France.

Licence: O. Dameron, Biostatistiques, 12h, 1st year school of medicine, Univ. Rennes 1, France

Licence: O. Dameron, C2i niveau 2, 2.5h, 2nd year school of medicine, Univ. Rennes 1, France

Licence: O. Dameron, Bioinformatique pour la génomique, 5h, 2nd year school of medicine, Univ. Rennes 1, France

Licence: C. Frioux, Programmation scientifique Python, 12h, L1, Rennes1, France.

Licence: C. Frioux, LaTeX, 12h, L3 ENSAI, France.

Licence: C. Frioux, Outils bureautiques pour le statisticien, 6h, L3 ENSAI, France.

Licence: C. Frioux, Algorithmique et programmation Python, 6h, L3 ENSAI, France.

Licence: L. Bourneuf, Ingénieurerie Systèmes et Réseaux, 10h, L3 INFO, France.

Licence: L. Bourneuf, Algorithmique des graphes, 8h, L3 INFO, France.

Licence: L. Bourneuf, Algorithmique des graphes, 2h, L3 MIAGE, France.

Master: L. Bourneuf, Principes de Programmation et d'Algorithmique, 6h, M1 BIG, France.

Master: L. Bourneuf, Projet, 10h, M1 BIG, France.

- Master: C. Belleannée, Programmation logique avec contraintes et algorithmes génétiques, 40h, M1 informatique, Univ. Rennes1, France.
- Master: C. Belleannée, Algorithmique du texte et bioinformatique, 10h, M1 informatique, Univ. Rennes1, France

Master: F. Coste, Apprentissage Automatique Supervisé, 10h, M2 Informatique, Univ. Rennes 1, France

Master: O. Dameron, Object-oriented programing, 20h, M1 bioinformatique et génomique, Univ. Rennes 1, France

Master: O. Dameron, Gestion de projet en informatique, 12h, M1 bioinformatique et génomique, Univ. Rennes 1, France

Master: O. Dameron, Ontologies biomédicales, 6h, Engineering school Institut Mines-Télécom Bretagne-Atlantique Brest, France

Master: O. Dameron, Internship jury, 25h, M1 bioinformatique et génomique, Univ. Rennes 1, France

Master: O. Dameron, Internship jury, 7.5h, M2 bioinformatique et génomique, Univ. Rennes 1, France

Master: O. Dameron, Intégration : remise à niveau en informatique, 14h, M1 bioinformatique, Univ. Rennes 1, France

Master: O. Dameron, Programmation impérative en Python, 39.5h, M1 bioinformatique, Univ. Rennes 1, France

Master: O. Dameron, Système informatique GNU/Linux, 12h, M1 bioinformatique, Univ. Rennes 1, France

Master: O. Dameron, Programmation en Python, 24h, M1 in Public Health, Univ. Rennes 1 [O. Dameron]

Master: O. Dameron, Semantic Web and bio-ontologies, 14h, M2 bioinformatique, Univ. Rennes 1, France

Master: O. Dameron, Bases de mathématiques et probabilités, 15h, M1 santé publique, Univ. Rennes 1, France

Master: A. Siegel, Integrative and Systems biology, 20h, M2, Univ. Rennes 1, France

Master: A. Siegel, Introduction to integrative biology, 2h, M2, Univ. Rennes 1, France

# 10.2.4. Supervision

PhD : Victorien Delannée, *Intégrer les échelles moléculaires et cellulaires dans l'inférence de réseaux métaboliques. Application aux xénobiotiques.*, started in Oct. 2014, defensed in Nov. 2017, supervised by A. Siegel and N. Théret [14].

PhD : Julie Laniau, *Structure de réseaux biologiques : rôle des noeuds internes vis-à-vis de la production de composés*, started in Oct. 2013, defended in Oct. 2017, supervised by A. Siegel and D. Eveillard. [15]

PhD : Jean Coquet, *Semantic-based reasoning for biological pathways analysis*, started in Oct. 2014, defended in Dec. 2017, supervised by O. Dameron and N. Théret [12]/

PhD in progress : Lucas Bourneuf, *Justifiable graph decomposition to assist biological network understanding*, started in Oct. 2016, supervised by J. Nicolas.

PhD in progress : Clémence Frioux, Using preferences in Answer Set Programming to decipher interactions within the species of an ecosystem at the genomic scale, started in Oct. 2015, supervised by A. Siegel.

PhD in progress : Yann Rivault, Analyse de parcours de soins à partir de bases de données médicoadministratives en utilisant des outils du Web Sémantique: identification de complications et de leurs déterminants suite à la pose chirurgicale de dispositif médical implantable en ambulatoire, started in Oct. 2015, supervised by O. Dameron and N. Lemeur.

PhD in progress : Juliette Talibart, *Learning grammars with long-distance correlations on proteins*, started in Nov. 2017, supervised by F. Coste and J. Nicolas.

PhD in progress : Mael Conan, *Predictive approach to assess the genotoxicity of environmental contaminants during liver fibrosis*, started in Oct. 2017, supervised by S. Langouet and A. Siegel.

PhD in progress: Marine Louarn, *Intégration de données génomiques massives et hétérogènes, application aux mutations non-codantes dans le lymphome folliculaire*, started in Oct. 2017, supervised by A. Siegel, T. Fest (CHU) and O. Dameron.

PhD in progress : Méline Wery, *Methodology development in disease treatment projects.*, started in Oct. 2017, supervised by O. Dameron, C. Bettembourg (Sanofi) and A. Siegel.

### 10.2.5. Juries

- Member of Ph-D thesis juries. J. Mercier, Univ. Evry/CEA [A. Siegel, reviewer]. C. Franay, INRA Toulouse [A. Siegel, reviewer]. W Bedhiafi, Univ. Pierre et Marie Curie Paris + UTM Tunis [O. Dameron]. V. Delannée, Univ. Rennes 1 [N. Theret, O. Dameron]. J. Coquet, Univ. Rennes 1 [O. Dameron, N. Theret]. P. Finet, Univ. Rennes 1 [O. Dameron]
- Member of habilitation thesis juries. E. Remy, Univ. Marseille [A. Siegel, president].
- Member of medicine doctorate juries G. Lebailly, Univ. Rennes 1 [O. Dameron].

### 10.2.6. Internships

- Internship, from Jun 2017 until Jul 2017. Supervised by J. Nicolas. Student: Alexis Baudin. Subject: Recherche d'attracteurs dans les réseaux booléens synchrones en ASP.
- Internship, from Jan until Jun 2017. Supervised by A. Siegel. Student: Mael Conan. Subject: Modélisation et caractérisation de la réponse au stress de la cyanobactérie marine Synechococcus sp. WH7803.
- Internship, from from Apr 2017 until Jul 2017. Supervised by Nathalie Théret and Olivier Dameron. Student: Kevin Courtet. Subject: Integration of genic regulatory interaction network by miRNAs from patients' macrophages with cystic fibrosis.
- Internship, from Apr 2017 until Jul 2017. Supervised by J. Got. Student: Nicolas Guillaudeux Subject: Vérifications du réseau métabolique entier de Tisochrysis lutea.
- Internship, from Apr 2017 until Jul 2017. Supervised by J. Nicolas and F. Coste. Student: AliHassan Kachalo Subject: Annotation automatique en familles des séquences d'une superfamille d'enzymes, les HAD (haloacides déhalogénases) par Analyse de Concepts Formels (FCA).
- Internship, from from Apr 2017 until Jul 2017. Supervised by C. Frioux and C. Trottier. Student: Claire Lippold. Subject: Exploration et caractérisation du microbiome associé à Ectocarpus subulatus str. BFT.
- Internship, from Jan 2017 until Jun 2017. Supervised by O. Dameron and A. Siegel. Student: Marine Louarn. Subject: Analysis and integration of heterogeneous large-scale genomics data.
- Internship, from Apr 2017 until Jul 2017. Supervised by . Théret and J. Nicolas. Student: Aurelie Nicolas. Subject: Modeling of interaction networks from extracellular matrix components using formal concept analyses.
- Internship, from Apr 2017 until Jul 2017. Supervised by C. Belleannée. Student: Dimitri Pedron. Subject: Annotation et prédiction de transcriptome: validation d'ORF alternatifs prédits. Application au gène CREM chez l'humain, la souris et le chien.
- Internship, from Feb 2017 until Jun 2017. Supervised by F. Coste. Student: Manon Ruffini, Subject: Better scoring schemes for the recognition of functional proteins by protomata

- Internship, from Feb 2017 until Jun 2017. Supervised by J. Nicolas. Student: MarieSalmon, Subject: Biclustering: quantitative formal concept analysis in Answer Set Programming
- Internship, from Jan 2017 until Jun 2017. Supervised by A. Siegel and O. Dameron. Student: Meline Wery Subject: Formalizing and computing signatures of phenotypes within a biological network
- Internship, from May 2017 until Jul 2017. Supervised by C. Belleannée. Student: Mohamed Zemmouri Subject: Analyse de texte en bioinformatique : Modélisation grammaticale d'un site ADN, et recherche du site, même dégénéré, sur l'intégralité du génome humain

# **10.3.** Popularization

# 10.3.1. General audience paper

We have written a contribution to the collaborative book edited by CNRS on the main issues of data-mining. Our contribution was specifically related to modeling issues arising in ecology with the development of NGS technologies [32].

### 10.3.2. Science en Cour[t]s

(*http://sciences-en-courts.fr/*) Many of our on-going and former Ph-D students (A. Antoine-Lorquin, C. Bettembourg, J. Coquet, V. Delannée, G. Garet, S. Prigent) have been heavily involved in organization of a local Popularization Festival where Ph.D. students explain their thesis via short movies. The movies are presented to a professional jury composed of artists and scientists, and of high-school students. Previous years films can be viewed on the festival web-site <sup>0</sup> In this context, the following scientific film has been created by the team members in 2017 *Helene et les cartons*<sup>0</sup>.

# 11. Bibliography

# Major publications by the team in recent years

- [1] C. BELLEANNÉE, O. SALLOU, J. NICOLAS. Logol: Expressive Pattern Matching in sequences. Application to Ribosomal Frameshift Modeling, in "PRIB2014 - Pattern Recognition in Bioinformatics, 9th IAPR International Conference", Stockholm, Sweden, M. COMIN, L. KALL, E. MARCHIORI, A. NGOM, J. RAJAPAKSE (editors), Springer International Publishing, August 2014, vol. 8626, p. 34-47 [DOI: 10.1007/978-3-319-09192-1\_4], https://hal.inria.fr/hal-01059506.
- [2] J. BOURDON, D. EVEILLARD, A. SIEGEL.Integrating quantitative knowledge into a qualitative gene regulatory network, in "PLoS Computational Biology", September 2011, vol. 7, n<sup>o</sup> 9 [DOI: 10.1371/JOURNAL.PCBI.1002157], http://hal.archives-ouvertes.fr/hal-00626708.
- [3] L. BOURNEUF, J. NICOLAS. FCA in a Logical Programming Setting for Visualization-oriented Graph Compression, in "International Conference on Formal Concept Analysis 2017", Rennes, France, Lecture Notes in Computer Sciences, June 2017, vol. 10308 [DOI: 10.1007/978-3-319-59271-8\_6], https://hal.archivesouvertes.fr/hal-01558302.
- [4] A. BRETAUDEAU, F. COSTE, F. HUMILY, L. GARCZAREK, G. LE CORGUILLÉ, C. SIX, M. RATIN, O. COLLIN, W. M. SCHLUCHTER, F. PARTENSKY. CyanoLyase: a database of phycobilin lyase sequences, motifs and functions, in "Nucleic Acids Research", November 2012, vol. 41 [DOI: 10.1093/NAR/GKS1091], http://hal.inria.fr/hal-00760946.

<sup>&</sup>lt;sup>0</sup>https://www.youtube.com/embed/3DRgLLITKUc

<sup>&</sup>lt;sup>0</sup>https://www.youtube.com/embed/yheO8Y0nWu0

- [5] F. COSTE, G. KERBELLEC. A Similar Fragments Merging Approach to Learn Automata on Proteins, in "ECML:Machine Learning: ECML 2005, 16th European Conference on Machine Learning, Porto, Portugal, October 3-7, 2005, Proceedings", J. GAMA, R. CAMACHO, P. BRAZDIL, A. JORGE, L. TORGO (editors), Lecture Notes in Computer Science, Springer, 2005, vol. 3720, p. 522-529.
- [6] M. GEBSER, C. GUZIOLOWSKI, M. IVANCHEV, T. SCHAUB, A. SIEGEL, P. VEBER, S. THIELE. Repair and Prediction (under Inconsistency) in Large Biological Networks with Answer Set Programming, in "Principles of Knowledge Representation and Reasoning", AAAI Press, 2010.
- [7] C. GUZIOLOWSKI, S. VIDELA, F. EDUATI, S. THIELE, T. COKELAER, A. SIEGEL, J. SAEZ-RODRIGUEZ. Exhaustively characterizing feasible logic models of a signaling network using Answer Set Programming, in "Bioinformatics", August 2013, vol. 29, n<sup>o</sup> 18, p. 2320-2326 [DOI: 10.1093/BIOINFORMATICS/BTT393], http://hal.inria.fr/hal-00853704.
- [8] S. PRIGENT, G. COLLET, S. M. DITTAMI, L. DELAGE, F. ETHIS DE CORNY, O. DAMERON, D. EVEILLARD, S. THIELE, J. CAMBEFORT, C. BOYEN, A. SIEGEL, T. TONON. *The genome-scale metabolic network of Ectocarpus siliculosus (EctoGEM): a resource to study brown algal physiology and beyond*, in "Plant Journal", September 2014, p. 367-81 [DOI: 10.1111/TPJ.12627], https://hal.archives-ouvertes.fr/hal-01057153.
- [9] S. PRIGENT, C. FRIOUX, S. M. DITTAMI, S. THIELE, A. LARHLIMI, G. COLLET, G. FABIEN, J. GOT, D. EVEILLARD, J. BOURDON, F. PLEWNIAK, T. TONON, A. SIEGEL. *Meneco, a Topology-Based Gap-Filling Tool Applicable to Degraded Genome-Wide Metabolic Networks*, in "PLoS Computational Biology", January 2017, vol. 13, n<sup>o</sup> 1, 32 [DOI: 10.1371/JOURNAL.PCBI.1005276], https://hal.inria.fr/hal-01449100.
- [10] C. ROUSSEAU, M. GONNET, M. LE ROMANCER, J. NICOLAS. CRISPI: a CRISPR interactive database, in "Bioinformatics", 2009, vol. 25, n<sup>o</sup> 24, p. 3317-3318.
- [11] S. VIDELA, J. SAEZ-RODRIGUEZ, C. GUZIOLOWSKI, A. SIEGEL.caspo: a toolbox for automated reasoning on the response of logical signaling networks families, in "Bioinformatics", 2017 [DOI: 10.1093/BIOINFORMATICS/BTW738], https://hal.inria.fr/hal-01426880.

# **Publications of the year**

### **Doctoral Dissertations and Habilitation Theses**

- [12] J. COQUET. Comprehensive study of large signaling pathways by clustering trajectories and characterization by semantic analysis, Université de Rennes 1, December 2017, https://hal.inria.fr/tel-01670730.
- [13] V. DELANNÉE. Integrate molecular and cellular scales in the inference of metabolic networks : application to *xenobiotics*, Université Rennes 1, November 2017, https://tel.archives-ouvertes.fr/tel-01690005.
- [14] V. DELANNÉE.*Integrate molecular and cellular scales in the inference of metabolic networks*. *Application to xenobiotics*, Univ. Rennes 1, November 2017, https://hal-univ-rennes1.archives-ouvertes.fr/tel-01659375.
- [15] J. LANIAU. *Biological networks structure : role of internal nodes regarding production of target elements*, Inria Rennes - Bretagne Atlantique, October 2017, https://hal.archives-ouvertes.fr/tel-01656474.

### **Articles in International Peer-Reviewed Journal**

- [16] M. P. CORTÉS, S. N. MENDOZA, D. TRAVISANY, A. GAETE, A. SIEGEL, V. CAMBIAZO, A. MAASS. Analysis of Piscirickettsia salmonis Metabolism Using Genome-Scale Reconstruction, Modeling, and Testing, in "Frontiers in Microbiology", December 2017, vol. 8, 15 [DOI: 10.3389/FMICB.2017.02462], https://hal.inria.fr/hal-01661270.
- [17] V. DELANNÉE, S. LANGOUËT, N. THÉRET, A. SIEGELA modeling approach to evaluate the balance between bioactivation and detoxification of MeIQx in human hepatocytes, in "PeerJ", 2017, vol. 5, e3703 [DOI: 10.7717/PEERJ.3703], https://hal.inria.fr/hal-01575579.
- [18] F. GONDRET, A. VINCENT, M. HOUÉE-BIGOT, A. SIEGEL, S. LAGARRIGUE, D. CAUSEUR, H. H. GILBERT, I. LOUVEAU.A transcriptome multi-tissue analysis identifies biological pathways and genes associated with variations in feed efficiency of growing pigs, in "BMC Genomics", 2017, vol. 18, n<sup>o</sup> 1, 244 [DOI: 10.1186/s12864-017-3639-0], https://hal.archives-ouvertes.fr/hal-01494107.
- [19] J. LANIAU, C. FRIOUX, J. NICOLAS, C. BAROUKH, M. P. CORTÉS, J. GOT, C. TROTTIER, D. EVEILLARD, A. SIEGEL. Combining graph and flux-based structures to decipher phenotypic essential metabolites within metabolic networks, in "PeerJ", 2017, vol. 5, e3860 [DOI : 10.7717/PEERJ.3860], https://hal.archivesouvertes.fr/hal-01635688.
- [20] V. PICARD, A. SIEGEL, J. BOURDON. A Logic for Checking the Probabilistic Steady-State Properties of Reaction Networks, in "Journal of computational biology : a journal of computational molecular cell biology", August 2017, vol. 24, n<sup>o</sup> 8, p. 1–12 [DOI : 10.1089/CMB.2017.0099], https://hal.archives-ouvertes.fr/hal-01552190.
- [21] S. PRIGENT, C. FRIOUX, S. M. DITTAMI, S. THIELE, A. LARHLIMI, G. COLLET, G. FABIEN, J. GOT, D. EVEILLARD, J. BOURDON, F. PLEWNIAK, T. TONON, A. SIEGEL. *Meneco, a Topology-Based Gap-Filling Tool Applicable to Degraded Genome-Wide Metabolic Networks*, in "PLoS Computational Biology", January 2017, vol. 13, n<sup>o</sup> 1, 32 [DOI: 10.1371/JOURNAL.PCBI.1005276], https://hal.inria.fr/hal-01449100.
- [22] S. VIDELA, J. SAEZ-RODRIGUEZ, C. GUZIOLOWSKI, A. SIEGEL.caspo: a toolbox for automated reasoning on the response of logical signaling networks families, in "Bioinformatics", 2017 [DOI: 10.1093/BIOINFORMATICS/BTW738], https://hal.inria.fr/hal-01426880.

### **Invited Conferences**

[23] V. WUCHER, F. LEGEAI, L. BOURNEUF, T. DERRIEN, A. GALLOT, S. HUDAVERDIAN, S. JAUBERT-POSSAMAI, N. LETERME-PRUNIER, J. NICOLAS, H. SEITZ, A. SIEGEL, S. TANGUY, G. LE TRION-NAIRE, D. TAGU. *Integrative genomics and gene networks for studying phenotypic plasticity in the pea aphid*, in "10th Arthropod Genomics Symposium", Notre Dame, United States, June 2017, https://hal.inria.fr/hal-01566438.

# **International Conferences with Proceedings**

[24] Best Paper

L. BOURNEUF, J. NICOLAS.*FCA in a Logical Programming Setting for Visualization-oriented Graph Compression*, in "International Conference on Formal Concept Analysis 2017", Rennes, France, Lecture Notes in Computer Sciences, June 2017, vol. 10308 [*DOI* : 10.1007/978-3-319-59271-8\_6], https://hal.archives-ouvertes.fr/hal-01558302.

[25] J. COQUET, N. THÉRET, V. LEGAGNEUX, O. DAMERON. Identifying Functional Families of Trajectories in Biological Pathways by Soft Clustering: Application to TGF-β Signaling, in "CMSB 2017 - 15th International Conference on Computational Methods in Systems Biology", Darmstadt, France, Lecture Notes in Computer Sciences, September 2017, 17, https://hal.archives-ouvertes.fr/hal-01559249.

[26] Best Paper

C. FRIOUX, T. SCHAUB, S. SCHELLHORN, A. SIEGEL, P. WANKO.*Hybrid Metabolic Network Completion*, in "14th International Conference on Logic Programming and Nonmonotonic Reasoning - LPNMR 2017", Espoo, Finland, M. BALDUCCINI, T. JANHUNEN (editors), Logic Programming and Nonmonotonic Reasoning, Springer, July 2017, vol. 10377, p. 308-321, https://hal.inria.fr/hal-01557347.

### **National Conferences with Proceeding**

[27] P. FINET, B. GIBAUD, O. DAMERON, R. LE BOUQUIN JEANNES. Interoperable infrastructure and implementation of a health data model for remote monitoring of chronic diseases with comorbidities, in "Journées d'Etude sur la TéléSANté, 6ème edition", Bourges, France, Pôle Capteurs, Université d'Orléans, May 2017, https://hal.archives-ouvertes.fr/hal-01565009.

### **Conferences without Proceedings**

- [28] X. GARNIER, A. BRETAUDEAU, O. FILANGI, F. LEGEAI, A. SIEGEL, O. DAMERON. AskOmics, a web tool to integrate and query biological data using semantic web technologies, in "JOBIM 2017 - Journées Ouvertes en Biologie, Informatique et Mathématiques", Lille, France, July 2017, 1, https://hal.inria.fr/hal-01577425.
- [29] X. GARNIER, O. DAMERON, O. FILANGI, F. LEGEAI, A. BRETAUDEAU. Integration of Linked Data into Galaxy using Askomics, in "Galaxy Community Conference", Montpellier, France, June 2017, https://hal. inria.fr/hal-01576870.
- [30] C. LIPPOLD, C. TROTTIER, C. FRIOUX, M. CHEVALLIER, S. M. DITTAMI, A. SIEGEL. *Toward the study of metabolic functions in algal holobionts*, in "4ème Colloque de Génomique Environnementale", Marseille, France, September 2017, https://hal.inria.fr/hal-01576484.
- [31] V. POTDEVIN, M. CHEVALLIER, R. GARRIDO OTER, A. SIEGEL, S. HACQUARD, D. EVEILLARD, P. VANDENKOORNHUYSE.*Toward a better understanding of the plant microbiota*, in "4ème colloque de Génomique Environnementale", Marseille, France, GDR Génomique Environnementale, September 2017, https://hal.inria.fr/hal-01576743.

### Scientific Books (or Scientific Book chapters)

[32] D. EVEILLARD, A. SIEGEL, P. VANDENKOORNHUYSE. *L'écologie des systèmes*, in "Les Big Data à Découvert", M. BOUZEGHOUB, R. MOSSERI (editors), CNRS éditions, 2017, https://hal.inria.fr/hal-01575603.

### **Other Publications**

[33] A. ANTOINE-LORQUIN, A. ARNAOTY, S. ASGARI, M. BATAILLER, L. BEAUCLAIR, C. BELLEANNÉE, S. BIRE, N. BUISINE, V. COUSTHAM, A. GIRAULT, S. GUYETANT, T. LECOMTE, B. PIÉGU, B. PITARD, I. STÉVANT, Y. BIGOT. *First landscape of binding to chromosomes for a domesticated mariner transposase in the human genome: diversity of genomic targets of SETMAR isoforms in two colorectal cell lines*, July 2017, working paper or preprint [DOI: 10.1101/115030], https://hal.archives-ouvertes.fr/hal-01558001.

- [34] A. ANTOINE-LORQUIN, F. MAHÉ, M. DUNTHORN, C. BELLEANNÉE. *Detection of mutated primers and impact on targeted metagenomics results*, July 2017, working paper or preprint, https://hal.archives-ouvertes. fr/hal-01564093.
- [35] C. MARCHET, L. LECOMPTE, C. DA SILVA, C. CRUAUD, J.-M. AURY, J. NICOLAS, P. PETERLONGO.De novo Clustering of Gene Expressed Variants in Transcriptomic Long Reads Data Sets, November 2017, working paper or preprint [DOI: 10.1101/170035], https://hal.archives-ouvertes.fr/hal-01643156.
- [36] M. RUFFINI.*Better scoring schemes for the recognition of functional proteins by protomata*, Rennes 1, June 2017, https://hal.inria.fr/hal-01557941.

# **References in notes**

- [37] O. ABDOU-ARBI, S. LEMOSQUET, J. VAN MILGEN, A. SIEGEL, J. BOURDON. Exploring metabolism flexibility in complex organisms through quantitative study of precursor sets for system outputs, in "BMC Systems Biology", 2014, vol. 8, n<sup>o</sup> 1, 8 [DOI: 10.1186/1752-0509-8-8], https://hal.inria.fr/hal-00947219.
- [38] V. ACUÑA, A. ARAVENA, C. GUZIOLOWSKI, D. EVEILLARD, A. SIEGEL, A. MAASS.Deciphering transcriptional regulations coordinating the response to environmental changes, in "BMC Bioinformatics", January 2016, vol. 17, n<sup>O</sup> 1, p. 129-42, https://hal.archives-ouvertes.fr/hal-01260866.
- [39] V. ACUÑA, A. ARAVENA, A. MAASS, A. SIEGEL.*Modeling parsimonious putative regulatory networks: complexity and heuristic approach*, in "15th conference in Verification, Model Checking, and Abstract Interpretation", San Diego, United States, Springer, 2014, vol. 8318, p. 322-336 [DOI : 10.1007/978-3-642-54013-4\_18], https://hal.inria.fr/hal-00926477.
- [40] G. ANDRIEUX, M. LE BORGNE, N. THÉRET.An integrative modeling framework reveals plasticity of TGF-Beta signaling, in "BMC Systems Biology", 2014, vol. 8, n<sup>o</sup> 1, 30 [DOI : 10.1186/1752-0509-8-30], http://www.hal.inserm.fr/inserm-00978313.
- [41] A. ANTOINE-LORQUIN, F. MAHÉ, M. DUNTHORN, C. BELLEANNÉE. Detection of mutated primers and impact on targeted metagenomics results, in "RCAM'16 "Recent Computational Advances in Metagenomics"", The Hague, Netherlands, September 2016, https://hal.inria.fr/hal-01576304.
- [42] A. ANTOINE-LORQUIN, F. MAHÉ, M. DUNTHORN, C. BELLEANNÉE.*Impact de la recherche d'amorces mutées sur les résultats d'analyses métagénomiques*, in "Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM)", Rennes, France, Jobim-2016, Société Française de Bioinformatique (SFBI), June 2016, https://hal.archives-ouvertes.fr/hal-01343121.
- [43] T. L. BAILEY, M. BODEN, F. A. BUSKE, M. FRITH, C. E. GRANT, L. CLEMENTI, J. REN, W. W. LI, W. S. NOBLE. *MEME Suite: tools for motif discovery and searching*, in "Nucleic Acids Research", 2009, vol. 37, n<sup>o</sup> suppl\_2, W202, http://dx.doi.org/10.1093/nar/gkp335.
- [44] T. L. BAILEY, J. JOHNSON, C. E. GRANT, W. S. NOBLE. *The MEME Suite*, in "Nucleic Acids Research", 2015, vol. 43, n<sup>o</sup> W1, p. W39-W49, http://dx.doi.org/10.1093/nar/gkv416.
- [45] F. BALUSSON, M.-A. BOTREL, O. DAMERON, Y. DAUXAIS, E. DREZEN, A. DUPUY, T. GUYET, D. GROSS-AMBLARD, A. HAPPE, N. LE MEUR, B. LE NAUTOUT, E. LERAY, E. NOWAK, C. RAULT,

E. OGER, E. POLARD.*PEPS: a platform for supporting studies in pharmaco-epidemiology using medico-administrative databases*, in "International Congress on e-Health Research", Paris, France, October 2016, https://hal.inria.fr/hal-01380939.

- [46] C. BARAL. Knowledge Representation, Reasoning and Declarative Problem Solving, Cambridge University Press, 2010.
- [47] T. BAUMURATOVA, D. SURDEZ, B. DELYON, G. STOLL, O. DELATTRE, O. RADULESCU, A. SIEGEL.Localizing potentially active post-transcriptional regulations in the Ewing's sarcoma gene regulatory network, in "BMC Systems Biology", 2010, vol. 4, n<sup>o</sup> 1, 146 [DOI: 10.1186/1752-0509-4-146], http://www.hal.inserm.fr/inserm-00984711.
- [48] R. BELLAZZI.*Big Data and Biomedical Informatics: A Challenging Opportunity*, in "Yearbook of medical informatics", 2014, vol. 9, n<sup>o</sup> 1, In press.
- [49] R. BELLAZZI, M. DIOMIDOUS, I. N. SARKAR, K. TAKABAYASHI, A. ZIEGLER, A. T. MCCRAY.*Data analysis and data mining: current issues in biomedical informatics*, in "Methods of information in medicine", 2011, vol. 50, n<sup>o</sup> 6, p. 536–544.
- [50] C. BELLEANNÉE, O. SALLOU, J. NICOLAS.Logol: Expressive Pattern Matching in sequences. Application to Ribosomal Frameshift Modeling, in "PRIB2014 - Pattern Recognition in Bioinformatics, 9th IAPR International Conference", Stockholm, Sweden, M. COMIN, L. KALL, E. MARCHIORI, A. NGOM, J. RAJAPAKSE (editors), Springer International Publishing, August 2014, vol. 8626, p. 34-47 [DOI: 10.1007/978-3-319-09192-1\_4], https://hal.inria.fr/hal-01059506.
- [51] R. BELLÉ, S. PRIGENT, A. SIEGEL, P. CORMIER. Model of cap-dependent translation initiation in sea urchin: a step towards the eukaryotic translation regulation network, in "Molecular Reproduction and Development", 2010, vol. 77, n<sup>o</sup> 3, p. 257-64.
- [52] R. BENDAOUD, A. NAPOLI, Y. TOUSSAINT. Formal Concept Analysis: A Unified Framework for Building and Refining Ontologies, in "Knowledge Engineering: Practice and Patterns, 16th International Conference, EKAW 2008, Acitrezza, Italy, September 29 - October 2, 2008. Proceedings", A. GANGEMI, J. EUZENAT (editors), Lecture Notes in Computer Science, Springer, 2008, vol. 5268, p. 156-171, http://dx.doi.org/10. 1007/978-3-540-87696-0\_16.
- [53] C. BETTEMBOURG, O. DAMERON, A. BRETAUDEAU, F. LEGEAI.AskOmics : IntA©gration et interrogation de réseaux de régulation génomique et post-génomique, in "IN OVIVE (INtégration de sources/masses de données hétérogÃ"nes et Ontologies, dans le domaine des sciences du VIVant et de lâEnvironnement)", Rennes, France, June 2015, 7, https://hal.inria.fr/hal-01184903.
- [54] C. BETTEMBOURG, C. DIOT, O. DAMERON. Semantic particularity measure for functional characterization of gene sets using gene ontology, in "PLoS ONE", 2014, vol. 9, n<sup>o</sup> 1, e86525 [DOI: 10.1371/JOURNAL.PONE.0086525], https://hal.inria.fr/hal-00941850.
- [55] C. BETTEMBOURG, C. DIOT, O. DAMERON. Optimal threshold determination for interpreting semantic similarity and particularity, in "PLoS ONE", 2015, vol. 10, n<sup>o</sup> 7, e0133579 [DOI: 10.1371/JOURNAL.PONE.0133579], https://hal.archives-ouvertes.fr/hal-01207763.

- [56] C. BETTEMBOURG, C. DIOT, O. DAMERON.Optimal Threshold Determination for Interpreting Semantic Similarity and Particularity: Application to the Comparison of Gene Sets and Metabolic Pathways Using GO and ChEBI, in "PLoS ONE", 2015, 30 [DOI: 10.1371/JOURNAL.PONE.0133579], https://hal.inria.fr/hal-01184934.
- [57] C. BIZER, T. HEATH, T. BERNERS LEE.*Linked Data–The story so far*, in "International Journal on Semantic Web and Information Systems", 2009, vol. 5, n<sup>o</sup> 3, p. 1–22.
- [58] J. A. BLAKE, C. J. BULT. Beyond the data deluge: Data integration and bio-ontologies, in "Journal of Biomedical Informatics", 2006, vol. 39, n<sup>o</sup> 3, p. 314–320.
- [59] P. BLAVY, F. GONDRET, S. LAGARRIGUE, J. VAN MILGEN, A. SIEGEL. Using a large-scale knowledge database on reactions and regulations to propose key upstream regulators of various sets of molecules participating in cell metabolism, in "BMC Systems Biology", 2014, vol. 8, n<sup>o</sup> 1, 32 [DOI: 10.1186/1752-0509-8-32], https://hal.inria.fr/hal-00980499.
- [60] P. BORDRON, D. EVEILLARD, A. MAASS, A. SIEGEL. An ASP application in integrative biology: identification of functional gene units, in "LPNMR - 12th Conference on Logic Programming and Nonmonotonic Reasoning - 2013", Corunna, Spain, September 2013, http://hal.inria.fr/hal-00853762.
- [61] P. BORDRON, M. LATORRE, M. P. CORTÉS, M. GONZALES, S. THIELE, A. SIEGEL, A. MAASS, D. EVEILLARD. Putative bacterial interactions from metagenomic knowledge with an integrative systems ecology approach, in "MicrobiologyOpen", 2015, vol. 5, n<sup>o</sup> 1, p. 106-117 [DOI: 10.1002/MB03.315], https://hal.inria.fr/hal-01246173.
- [62] A. BRETAUDEAU, F. COSTE, F. HUMILY, L. GARCZAREK, G. LE CORGUILLÉ, C. SIX, M. RATIN, O. COLLIN, W. M. SCHLUCHTER, F. PARTENSKY. CyanoLyase: a database of phycobilin lyase sequences, motifs and functions, in "Nucleic Acids Research", November 2012, 6 [DOI : 10.1093/NAR/GKS1091], https://hal.inria.fr/hal-01094087.
- [63] H. CHEN, T. YU, J. Y. CHEN. Semantic Web meets Integrative Biology: a survey, in "Briefings in bioinformatics", 2012, vol. 14, n<sup>o</sup> 1, p. 109–125.
- [64] M. CHEVALLIER, M. AITE, J. GOT, G. COLLET, N. LOIRA, M. P. CORTÉS, C. FRIOUX, J. LANIAU, C. TROTTIER, A. MAASS, A. SIEGEL.*Handling the heterogeneity of genomic and metabolic networks data within flexible workflows with the PADMet toolbox*, in "Jobim 2016 : 17ème Journées Ouvertes en Biologie, Informatique et Mathématiques", Lyon, France, June 2016, https://hal.inria.fr/hal-01377844.
- [65] G. COLLET, D. EVEILLARD, M. GEBSER, S. PRIGENT, T. SCHAUB, A. SIEGEL, S. THIELE.*Extending the Metabolic Network of Ectocarpus Siliculosus using Answer Set Programming*, in "LPNMR 12th Conference on Logic Programming and Nonmonotonic Reasoning 2013", Corunna, Spain, September 2013, http://hal. inria.fr/hal-00853752.
- [66] F. COSTE.Learning the Language of Biological Sequences, in "Topics in Grammatical Inference", J. HEINZ, J. M. SEMPERE (editors), Springer-Verlag, 2016 [DOI: 10.1007/978-3-662-48395-4\_8], https://hal.inria. fr/hal-01244770.

- [67] F. COSTE, G. GARET, A. GROISILLIER, J. NICOLAS, T. TONON. Automated Enzyme classification by Formal Concept Analysis, in "ICFCA - 12th International Conference on Formal Concept Analysis", Cluj-Napoca, Romania, Springer, June 2014, https://hal.inria.fr/hal-01063727.
- [68] S. DAMINELLI, V. J. HAUPT, M. REIMANN, M. SCHROEDER. Drug repositioning through incomplete bicliques in an integrated drug-target-disease network, in "Integr. Biol.", 2012, vol. 4, p. 778-788, http://dx.doi. org/10.1039/C2IB00154C.
- [69] S. M. DITTAMI, T. BARBEYRON, C. BOYEN, J. CAMBEFORT, G. COLLET, L. DELAGE, A. GOBET, A. GROISILLIER, C. LEBLANC, G. MICHEL, D. SCORNET, A. SIEGEL, J. E. TAPIA, T. TONON.Genome and metabolic network of "Candidatus Phaeomarinobacter ectocarpi" Ec32, a new candidate genus of Alphaproteobacteria frequently associated with brown algae, in "Frontiers in Genetics", 2014, vol. 5, 241 [DOI: 10.3389/FGENE.2014.00241], https://hal.inria.fr/hal-01079739.
- [70] A. EVRARD, C. BETTEMBOURG, M. JUBAULT, O. DAMERON, O. FILANGI, A. BRETAUDEAU, F. F. LEGEAI. *Integration and query of biological datasets with Semantic Web technologies: AskOmics*, June 2016, Journées Ouvertes Biologie, Informatique et Mathématiques (JOBIM 2016), Poster, https://hal.inria.fr/hal-01391087.
- [71] R. D. FINN, J. CLEMENTS, W. ARNDT, B. L. MILLER, T. J. WHEELER, F. SCHREIBER, A. BATEMAN, S. R. EDDY.*HMMER web server: 2015 update*, in "Nucleic Acids Research", 2015, vol. 43, n<sup>o</sup> W1, W30, http://dx.doi.org/10.1093/nar/gkv397.
- [72] C. GALIEZ, M. CHRISTOPHE, F. COSTE, P. BALDI.VIRALpro: a tool to identify viral capsid and tail sequences, in "Bioinformatics", 2016 [DOI : 10.1093/BIOINFORMATICS/BTV727], https://hal.archivesouvertes.fr/hal-01242251.
- [73] M. GEBSER, R. KAMINSKI, B. KAUFMANN, T. SCHAUB. Answer Set Solving in Practice, Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan and Claypool Publishers, 2012.
- [74] F. GONDRET, I. LOUVEAU, M. HOUEE, D. CAUSEUR, A. SIEGEL.*Data integration*, in "Meeting INRA-ISU", Ames, United States, March 2015, 11, https://hal.archives-ouvertes.fr/hal-01210940.
- [75] F. GONDRET, A. VINCENT, M. HOUEE, S. LAGARRIGUE, A. SIEGEL, D. CAUSEUR, I. LOU-VEAU.Integrative responses of pig adipose tissues to high-fat high-fiber diet: towards key regulators of energy flexibility, March 2015, ASAS/ADSA midwest meeting, Poster, https://hal.archives-ouvertes.fr/hal-01210925.
- [76] F. GONDRET, A. VINCENT, M. HOUÉE-BIGOT, A. SIEGEL, S. LAGARRIGUE, I. LOUVEAU, D. CAUSEUR. Molecular alterations induced by a high-fat high-fiber diet in porcine adipose tissues: variations according to the anatomical fat location, in "BMC Genomics", December 2016, vol. 17, n<sup>o</sup> 1, 120, article number 120 [DOI: 10.1186/s12864-016-2438-3], https://hal.archives-ouvertes.fr/hal-01286555.
- [77] C. GUZIOLOWSKI, A. BOURDÉ, F. MOREEWS, A. SIEGEL.BioQuali Cytoscape plugin: analysing the global consistency of regulatory networks, in "Bmc Genomics", 2009, vol. 26, n<sup>o</sup> 10, 244 [DOI: 10.1186/1471-2164-10-244], http://hal.inria.fr/inria-00429804.
- [78] C. GUZIOLOWSKI, S. VIDELA, F. EDUATI, S. THIELE, T. COKELAER, A. S. SIEGEL, J. SAEZ-RODRIGUEZ. Exhaustively characterizing feasible logic models of a signaling network using Answer Set Pro-

*gramming*, in "Bioinformatics", 2013, vol. 29, p. 2320-2326 [*DOI* : 10.1093/BIOINFORMATICS/BTT393], https://hal.inria.fr/hal-00853704v1.

- [79] F. HERAULT, A. VINCENT, O. DAMERON, P. LE ROY, P. CHEREL, M. DAMON. *The longissimus and semimembranosus muscles display marked differences in their gene expression profiles in pig*, in "PLoS ONE", 2014, vol. 9, n<sup>o</sup> 5, e96491 [DOI : 10.1371/JOURNAL.PONE.0096491], https://hal.inria.fr/hal-00989635.
- [80] R. KAMINSKI, A. SIEGEL, T. SCHAUB, S. VIDELA. *Minimal Intervention Strategies in Logical Signaling Networks with ASP*, in "Theory and Practice of Logic Programming", September 2013, vol. 13, n<sup>o</sup> Special issue 4-5, p. 675-690 [DOI: 10.1017/S1471068413000422], http://hal.inria.fr/hal-00853747.
- [81] H. KLARNER, A. STRECK, H. SIEBERT.PyBoolNet: a python package for the generation, analysis and visualization of boolean networks, in "Bioinformatics", 2017, vol. 33, n<sup>o</sup> 5, p. 770-772, http://dx.doi.org/ 10.1093/bioinformatics/btw682.
- [82] S. KURTZ. *The Vmatch large scale sequence analysis software*, in "Ref Type: Computer Program", 2003, vol. 412.
- [83] S. LAURENT, A. RICHARD, O. MULNER-LORILLON, J. MORALES, D. FLAMENT, V. GLIPPA, J. BOUR-DON, P. GOSSELIN, A. SIEGEL, P. CORMIER, R. BELLÉ. Modelization of the regulation of protein synthesis following fertilization in sea urchin shows requirement of two processes: a destabilization of eIF4E:4E-BP complex and a great stimulation of the 4E-BP-degradation mechanism, both rapamycin-sensitive, in "Frontiers in Genetics", 2014, vol. 5, 117 [DOI: 10.3389/FGENE.2014.00117], https://hal.inria.fr/hal-01079758.
- [84] F. LEGEAI, C. BETTEMBOURG, A. BRETAUDEAU, Y. CHAUSSIN, O. DAMERON, D. TAGU.BIPAA/Askomics, a new and easy approach for querying genomics and epigenomics elements in interaction, in "XXVth International Congress of Entomology 2016", Orlando, Florida, United States, September 2016, https://hal.inria.fr/hal-01391080.
- [85] F. LEGEAI, T. DERRIEN, V. WUCHER, D. AUDREY, G. LE TRIONNAIRE, D. TAGU.Long non-coding RNA in the pea aphid; identification and comparative expression in sexual and asexual embryos, in "Arthropod Genomics Symposium", Urbana, United States, June 2014, https://hal.inria.fr/hal-01091304.
- [86] S.-W. LEUNG, C. MELLISH, D. ROBERTSON.Basic Gene Grammars and DNA-ChartParser for language processing of Escherichia coli promoter DNA sequences, in "Bioinformatics", 2001, vol. 17, n<sup>o</sup> 3, p. 226-236 [DOI : 10.1093/BIOINFORMATICS/17.3.226], http://bioinformatics.oxfordjournals.org/content/17/ 3/226.abstract.
- [87] A. LIHONOSOVA, A. KAMINSKAYA. Using Formal Concept Analysis for Finding the Closest Relatives among a Group of Organisms, in "Procedia Computer Science", 2014, vol. 31, n<sup>o</sup> Complete, p. 860-868, http://dx. doi.org/10.1016/j.procs.2014.05.337.
- [88] K. M. LIVINGSTON, M. BADA, W. A. BAUMGARTNER, L. E. HUNTER. *KaBOB: ontology-based semantic integration of biomedical databases*, in "BMC bioinformatics", 2015, vol. 16, 126.
- [89] M. MARTIN. Cutadapt removes adapter sequences from high-throughput sequencing reads, in "EMBnet.journal", 2011, vol. 17, n<sup>o</sup> 1, 10, http://dx.doi.org/10.14806/ej.17.1.200.

- [90] N. MESSAI, M.-D. DEVIGNES, A. NAPOLI, M. SMAIL-TABBONE.*Many-Valued Concept Lattices for Conceptual Clustering and Information Retrieval*, in "Proceedings of the 2008 Conference on ECAI 2008: 18th European Conference on Artificial Intelligence", Amsterdam, The Netherlands, The Netherlands, IOS Press, 2008, p. 127–131, http://dl.acm.org/citation.cfm?id=1567281.1567313.
- [91] P. T. MONTEIRO, C. CHAOUIYA. Efficient Verification for Logical Models of Regulatory Networks, M. P. ROCHA, N. LUSCOMBE, F. FDEZ-RIVEROLA, J. M. C. RODRÍGUEZ (editors), Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, p. 259–267, https://doi.org/10.1007/978-3-642-28839-5\_30.
- [92] J. NICOLAS, P. DURAND, G. RANCHY, S. TEMPEL, A.-S. VALIN.Suffix-Tree Analyser (STAN): looking for nucleotidic and peptidic patterns in genomes, in "Bioinformatics (Oxford, England)", 2005, vol. 21(24), p. 4408-4410, https://hal.archives-ouvertes.fr/hal-00015234.
- [93] M. OBITKO, V. SNÁSEL, J. SMID.Ontology Design with Formal Concept Analysis, in "Proceedings of the CLA 2004 International Workshop on Concept Lattices and their Applications, Ostrava, Czech Republic, September 23-24, 2004", V. SNÁSEL, R. BELOHLÁVEK (editors), CEUR Workshop Proceedings, CEUR-WS.org, 2004, vol. 110, http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-110/paper12. pdf.
- [94] M. OSTROWSKI, L. PAULEVÉ, T. SCHAUB, A. SIEGEL, C. GUZIOLOWSKI. Boolean Network Identification from Perturbation Time Series Data combining Dynamics Abstraction and Logic Programming, in "BioSystems", 2016 [DOI: 10.1016/J.BIOSYSTEMS.2016.07.009], https://hal.archives-ouvertes.fr/hal-01354075.
- [95] V. PICARD, O. MULNER-LORILLON, J. BOURDON, J. MORALES, P. CORMIER, A. SIEGEL, R. BELLÉ. Model of the delayed translation of cyclin B maternal mRNA after sea urchin fertilization, in "Molecular Reproduction and Development", 2016 [DOI: 10.1002/MRD.22746], http://hal.upmc.fr/hal-01390047.
- [96] P.-F. PLUCHON, T. FOUQUEAU, C. CREZE, S. LAURENT, J. BRIFFOTAUX, G. HOGREL, A. PALUD, G. HENNEKE, A. GODFROY, W. HAUSNER, M. THOMM, J. NICOLAS, D. FLAMENT. An Extended Network of Genomic Maintenance in the Archaeon Pyrococcus abyssi Highlights Unexpected Associations between Eucaryotic Homologs, in "PLoS ONE", 2013, vol. 8, n<sup>o</sup> 11, e79707 [DOI: 10.1371/JOURNAL.PONE.0079707], http://hal.inria.fr/hal-00911795.
- [97] S. PRIGENT, G. COLLET, S. M. DITTAMI, L. DELAGE, F. ETHIS DE CORNY, O. DAMERON, D. EVEIL-LARD, S. THIELE, J. CAMBEFORT, C. BOYEN, A. SIEGEL, T. TONON. *The genome-scale metabolic network* of Ectocarpus siliculosus (EctoGEM): a resource to study brown algal physiology and beyond, in "Plant Journal", September 2014, p. 367-81 [DOI: 10.1111/TPJ.12627], https://hal.archives-ouvertes.fr/hal-01057153.
- [98] Y. RIVAULT, N. LE MEUR, O. DAMERON. Complications post-opératoires et mode de prise en charge en angioplastie : apport du Programme de Médicalisation des Systèmes d'Information (PMSI), in "Congrès Adelf-Epiter 2016", Rennes, France, Congrès Adelf-Epiter 2016, September 2016, https://hal.inria.fr/hal-01389688.
- [99] Y. RIVAULT, N. LE MEUR, O. DAMERON.La gestion de données médico-administratives grâce aux outils du Web sémantique, in "Journées ADELF-EMOIS " Système d'information hospitalier et Epidémiologie "", Dijon, France, Revue d'Épidémiologie et de Santé Publique, March 2016, vol. 64, n<sup>o</sup> Supplement 1, S15 [DOI: 10.1016/J.RESPE.2016.01.051], https://hal.archives-ouvertes.fr/hal-01297389.

- [100] A. ROSIER, P. MABO, L. TEMAL, P. VAN HILLE, O. DAMERON, L. DELÉGER, C. GROUIN, P. ZWEIGENBAUM, J. JACQUES, E. CHAZARD, L. LAPORTE, C. HENRY, A. BURGUN. Personalized and automated remote monitoring of atrial fibrillation, in "EP-Europace", 2016, vol. 18, n<sup>o</sup> 3, p. 347-352 [DOI: 10.1093/EUROPACE/EUV234], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01331019.
- [101] A. ROSIER, P. MABO, L. TEMAL, P. VAN HILLE, O. DAMERON, L. DELÉGER, C. GROUIN, P. ZWEIGEN-BAUM, J. JACQUES, E. CHAZARD, L. LAPORTE, C. HENRY, A. BURGUN. *Remote Monitoring of Cardiac Implantable Devices: Ontology Driven Classification of the Alerts*, in "Studies in Health Technology and Informatics", 2016, vol. 221, p. 59–63 [DOI : 10.3233/978-1-61499-633-0-59], https://hal-univ-rennes1. archives-ouvertes.fr/hal-01319949.
- [102] M. ROUANE-HACENE, M. HUCHARD, A. NAPOLI, P. VALTCHEV. Using Formal Concept Analysis for Discovering Knowledge Patterns, in "CLA'10: 7th International Conference on Concept Lattices and Their Applications", Sevilla, Spain, M. KRYSZKIEWICZ, S. OBIEDKOV (editors), CEUR, University of Sevilla, October 2010, n<sup>o</sup> 672, p. 223-234, http://hal-lirmm.ccsd.cnrs.fr/lirmm-00531802.
- [103] C. ROUSSEAU, M. GONNET, M. LE ROMANCER, N. NICOLAS. CRISPI: a CRISPR Interactive database, in "Bioinformatics", 2009, vol. 25, p. 3317-3318 [DOI: 10.1093/BIOINFORMATICS/BTP586], https://hal.inria. fr/inria-00438512.
- [104] L. ROYER, M. REIMANN, B. ANDREOPOULOS, M. SCHROEDER. Unraveling Protein Networks with Power Graph Analysis, in "PLoS Comput Biol", 07 2008, vol. 4, n<sup>o</sup> 7, e1000108, http://dx.plos.org/10. 1371%2Fjournal.pcbi.1000108.
- [105] M. RUFFINI. Better scoring schemes for the recognition of functional proteins by protomata, Rennes 1, June 2017, https://hal.inria.fr/hal-01557941.
- [106] T. SCHAUB, A. SIEGEL, S. VIDELA. Reasoning on the response of logical signaling networks with Answer Set Programming, in "Logical Modeling of Biological Systems", Wiley Online Librairy, 2014, p. 49-92 [DOI: 10.1002/9781119005223.CH2], https://hal.inria.fr/hal-01079762.
- [107] T. SCHAUB, S. THIELE. *Metabolic Network Expansion with Answer Set Programming*, in "ICLP 2009", LNCS, Springer, 2009, vol. 5649, p. 312-326.
- [108] S. SCHULZ, L. BALKANYI, R. CORNET, O. BODENREIDER. From Concept Representations to Ontologies: A Paradigm Shift in Health Informatics?, in "Healthcare informatics research", 2013, vol. 19, n<sup>o</sup> 4, p. 235-242.
- [109] D. SEARLS. The language of genes, in "Nature", 2002, vol. 420, p. 211-217.
- [110] D. SEARLS.String Variable Grammar: A Logic Grammar Formalism for the Biological Language of DNA, in "Journal of Logic Programming", 1995, vol. 24, n<sup>o</sup> 1&2, p. 73-102.
- [111] C. J. A. SIGRIST, E. DE CASTRO, L. CERUTTI, B. A. CUCHE, N. HULO, A. BRIDGE, L. BOUGUELERET, I. XENARIOS.*New and continuing developments at PROSITE*, in "Nucleic Acids Research", 2013, vol. 41, n<sup>o</sup> D1, D344, http://dx.doi.org/10.1093/nar/gks1067.
- [112] Z. D. STEPHENS, S. Y. LEE, F. FAGHRI, R. H. CAMPBELL, C. ZHAI, M. J. EFRON, R. IYER, M. C. SCHATZ, S. SINHA, G. E. ROBINSON. *Big Data: Astronomical or Genomical?*, in "PLoS biology", 2015, vol. 13, n<sup>0</sup> 7, e1002195.

- [113] S. TEMPEL, C. ROUSSEAU, F. TAHI, J. NICOLAS. ModuleOrganizer: detecting modules in families of transposable elements, in "BMC Bioinformatics", 2010, vol. 11, 474 [DOI: 10.1186/1471-2105-11-474], http://hal.inria.fr/inria-00536742.
- [114] S. THIELE, J. SAEZ-RODRIGUEZ, L. CERONE, A. SIEGEL, C. GUZIOLOWSKI, S. KLAMT. Extended notions of sign consistency to relate experimental data to signaling and regulatory network topologies, in "BMC Bioinformatics", 2015, vol. 16, 345 [DOI : 10.1186/s12859-015-0733-7], https://hal.inria.fr/hal-01225228.
- [115] S. VIDELA, C. GUZIOLOWSKI, F. EDUATI, S. THIELE, M. GEBSER, J. NICOLAS, J. SAEZ-RODRIGUEZ, T. SCHAUB, A. SIEGEL. *Learning Boolean logic models of signaling networks with ASP*, in "Journal of Theoretical Computer Science (TCS)", September 2015, n<sup>o</sup> 599, p. 79-101 [DOI: 10.1016/J.TCS.2014.06.022], https://hal.inria.fr/hal-01058610.
- [116] S. VIDELA, I. KONOKOTINA, L. ALEXOPOULOS, J. SAEZ-RODRIGUEZ, T. SCHAUB, A. SIEGEL, C. GUZIOLOWSKI. Designing experiments to discriminate families of logic models, in "Frontiers in Bioengineering and Biotechnology", September 2015, 9 [DOI : 10.3389/FBIOE.2015.00131], https://hal.inria.fr/hal-01196178.
- [117] R. WILLE.*restructuring lattice theory: an approach based on hierarchies of concepts*, in "Proceedings of the 7th International Conference on Formal Concept Analysis", Berlin, Heidelberg, ICFCA '09, Springer-Verlag, 2009, p. 314–339, http://dx.doi.org/10.1007/978-3-642-01815-2\_23.
- [118] V. WUCHER, D. TAGU, J. NICOLAS., B. LAUSEN, S. KROLAK-SCHWERDT, M. BÖHMER (editors)Edge Selection in a Noisy Graph by Concept Analysis – Application to a Genomic Network, Data Science, Learning by Latent Structures, and Knowledge Discovery, Springer, 2014, 550, https://hal.inria.fr/hal-01093337.
- [119] V. WUCHER. Modeling of a gene network between mRNAs and miRNAs to predict gene functions involved in phenotypic plasticity in the pea aphid, Université Rennes 1, November 2014, https://tel.archives-ouvertes.fr/ tel-01135870.
- [120] V. WUCHER. Modeling of a gene network between mRNAs and miRNAs to predict gene functions involved in phenotypic plasticity in the pea aphid, Universite Rennes 1, November 2014, https://hal.archives-ouvertes.fr/ tel-01095967.
- [121] T. YAN, D. YOO, T. Z. BERARDINI, L. A. MUELLER, D. C. WEEMS, S. WENG, J. M. CHERRY, S. Y. RHEE.*PatMatch: a program for finding patterns in peptide and nucleotide sequences*, in "Nucleic Acids Research", 2005, vol. 33, n<sup>o</sup> Web-Server-Issue, p. 262–266, https://doi.org/10.1093/nar/gki368.

# **Project-Team FLUMINANCE**

# Fluid Flow Analysis, Description and Control from Image Sequences

IN COLLABORATION WITH: Institut de recherche mathématique de Rennes (IRMAR)

IN PARTNERSHIP WITH: Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Earth, Environmental and Energy Sciences

# **Table of contents**

1.	Perso	nnel	309
2.	Overall Objectives		
3.	Research Program		
	3.1.	Estimation of fluid characteristic features from images	311
	3.2.	Data assimilation and Tracking of characteristic fluid features	311
	3.3.	Optimization and control of fluid flows with visual servoing	312
	3.4.	Numerical models applied to hydrogeology and geophysics	313
	3.5.	Numerical algorithms and high performance computing	313
4.	Appli		313
	4.1.	Introduction	313
	4.2.	Environmental sciences	314
-	4.3.	Experimental fluid mechanics and industrial flows	314
5.	Highi	ights of the year	
6.	New 3		314
	6.1.	2DLayered Motion	314
	6.2.	3DLayered Motion	315
	6.3.	Densemotion	315
	0.4.		210
	0.3.	I I PHOUN	216
	0.0.	H2OLAD DALMTDEE	216
	0.7. 6.9		217
7	0.0. Now 1		317
7.	7 1	Eluid motion estimation	
	7.1.	1 Stochastic uncertainty models for motion estimation	317
	7.1	2. Surface Currents estimation from Shore Based Videos	317
	7.1	2. Development of an image based measurement method for large scale characteriza	tion of
	/.1	indoor airflows	318
	7 1	A 3D flows reconstruction from image data	318
	7.1	5 Sparse representation algorithms	310
	7.1	Tracking Data assimilation and model-data coupling	319
	7.2.	Determined on the formation and model data coupling of large scale dynamical systems and	image
	1.2	data	319
	7 2	2 Ensemble variational data assimilation of large-scale dynamics with uncertainty	319
	7.2	<ul> <li>Reduced-order models for flows representation from image data</li> </ul>	320
	73	Analysis and modeling of turbulent flows and geophysical flows	320
	7 3	Geophysical flows modeling under location uncertainty	320
	7 3	2. Large eddies simulation models under location uncertainty	321
	7.3	3. Singular and regular solutions to the Navier-Stokes equations (NSE) and relative tu	rbulent
	, 10	models	321
	7.3	.4. Turbulence similarity theory for the modeling of Ocean Atmosphere interface	322
	7.3	.5. Hot-wire anemometry at low velocities	322
	7.3	.6. Numerical and experimental image and flow database	322
	7.4.	Visual servoing approach for fluid flow control	323
	7.5.	Coupled models in hydrogeology	323
	7.5	.1. Coupling of subsurface and seepage flows	323
	7.5	2. Characterizations of Solutions in Geochemistry	323
	7.5	.3. Reactive transport in fractured-porous media	324
	7.5	.4. Reactive transfers for multi-phasic flows	324

	7.6. Linear solvers	324
	7.6.1. Variable s-step GMRES	324
	7.6.2. Krylov method applied to reactive transport	324
8.	Bilateral Contracts and Grants with Industry	324
	8.1.1. Contract CERSAT/IFREMER	324
	8.1.2. Contract inter Carno IFREMER Inria	325
	8.1.3. Contract ITGA	325
	8.1.4. Contract CSTB	325
	8.1.5. ANDRA project	325
	8.1.6. IFPEN project	325
9.	Partnerships and Cooperations	326
	9.1. National Initiatives	326
	9.1.1. Comins'lab: SEACS : Stochastic modEl-dAta-Coupled representationS for the ana	ılysis,
	simulation and reconstruction of upper ocean dynamics	326
	9.1.2. ANR JCJC GERONIMO : Advanced GEophysical Reduced-Order Model constru	uction
	from IMage Observations	326
	9.1.3. ANR BECOSE : Beyond Compressive Sensing: Sparse approximation algorithms f	or ill-
	conditioned inverse problems.	326
	9.1.4. ANR-MN: H2MNO4 project	326
	9.1.5. GDR MANU	327
	9.2. International Initiatives	327
	9.2.1. Inria Associate Teams Not Involved in an Inria International Labs	327
	9.2.1.1. LFD-FLU	327
	9.2.1.2. Informal International Partners	327
	9.2.2. Participation in Other International Programs	327
10	9.3. International Research Visitors	327
10.		328
	10.1.1. Promoting Scientific Activities	328
	10.1.1. Scientific Events Organisation	328
	10.1.1.1. General Chair, Scientific Chair	328
	10.1.1.2. Member of the Organizing Committees	328
	10.1.2. Journal	320
	10.1.3. Journal	320
	10.1.3.1. Method of the Editorial Dodids	320
	10.1.3.2. Reviewer - Reviewing Activities	320
	10.1.5 Leadership within the Scientific Community	329
	10.1.6 Scientific Expertise	329
	10.1.7 Research Administration	329
	10.2. Teaching - Supervision - Juries	330
	10.2.1. Teaching	330
	10.2.2. Supervision	330
	10.2.3. Juries	330
	10.3. Popularization	331
11.	Bibliography	331

# **Project-Team FLUMINANCE**

Creation of the Project-Team: 2009 July 01

# **Keywords:**

# **Computer Science and Digital Science:**

- A3. Data and knowledge
- A3.3. Data and knowledge analysis
- A3.4. Machine learning and statistics
- A5.3. Image processing and analysis
- A5.4. Computer vision
- A5.9. Signal processing
- A6. Modeling, simulation and control
- A6.1. Mathematical Modeling
- A6.1.2. Stochastic Modeling (SPDE, SDE)
- A6.1.4. Multiscale modeling
- A6.2. Scientific Computing, Numerical Analysis & Optimization
- A6.2.1. Numerical analysis of PDE and ODE
- A6.2.7. High performance computing
- A6.3. Computation-data interaction
- A6.3.1. Inverse problems
- A6.3.2. Data assimilation
- A6.3.3. Data processing
- A6.3.4. Model reduction
- A6.3.5. Uncertainty Quantification
- A6.4. Automatic control

# **Other Research Topics and Application Domains:**

- B3.2. Climate and meteorology
- B3.3. Geosciences
- B5. Industry of the future
- B5.2. Design and manufacturing

# 1. Personnel

### **Research Scientists**

Etienne Mémin [Team leader, Inria, Senior Researcher, HDR] Jocelyne Erhel [Inria, Senior Researcher, HDR] Christophe Collewet [IRSTEA, Researcher, HDR] Dominique Heitz [IRSTEA, Researcher] Cédric Herzet [Inria, Researcher]

### **Faculty Member**

Roger Lewandowski [Univ de Rennes I, Professor, HDR]

## **Technical Staff**

Yvan Crenner [Inria, until Sep 2017]

#### **PhD Students**

Mohamed Yacine Ben Ali [CSTB, from Oct 2017] Pranav Chandramouli [Inria] Bastien Hamlat [Institut Français du Pétrole Energies Nouvelles] Musaab Khalid Osman Mohammed [IRSTEA] Long Li [Univ de Rennes I, from Oct 2017] Benoit Pinier [Univ de Rennes I, until Sep 2017] Romain Schuster [ITGA]

#### **Post-Doctoral Fellows**

Carlo Cintolesi [Inria, from Nov 2017] Pierre Derian [Inria, until Aug 2017] Mamadou Diallo [Inria, until Sep 2017] Thibaut Tronchin [Inria, until Oct 2017]

### Visiting Scientist

Shengze Cai [Zhejiang University, until Nov 2017]

### Administrative Assistant

Huguette Béchu [Inria]

# 2. Overall Objectives

# 2.1. Overall Objectives

The research group that we have entitled FLUMINANCE from a contraction between the words "Fluid" and "Luminance" is dedicated to the extraction of information on fluid flows from image sequences and to the development of tools for the analysis and control of these flows. The objectives of the group are at the frontiers of several important domains that range from fluid mechanics to geophysics. One of the main originality of the FLUMINANCE group is to combine cutting-edge researches on data-assimilation and flow numerical modeling with an ability to conduct proper intensive experimental validations on prototype flows mastered in laboratory. The scientific objectives decompose in four main themes:

#### • Fluid flows characterization from images

In this first axis, we aim at providing accurate measurements and consistent analysis of complex fluid flows through image analysis techniques. The application domain ranges from industrial processes and experimental fluid mechanics to environmental sciences. This theme includes also the use of non-conventional imaging techniques such as Schlieren techniques, Shadowgraphs, holography. The objective will be here to go towards 3D dense velocity measurements.

### Coupling dynamical model and image data

We focus here on the study, through image data, of complex and partially known fluid flows involving complex boundary conditions, multi-phase fluids, fluids and structures interaction problems. Our credo is that image analysis can provide sufficiently fine observations on small and medium scales to construct models which, applied at medium and large scale, account accurately for a wider range of the dynamics scales. The image data and a sound modeling of the dynamical uncertainty at the observation scale should allow us to reconstruct the observed flow and to provide efficient real flows (experimental or natural) based dynamical modeling. Our final goal will be to go towards a 3D reconstruction of real flows, or to operate large motion scales simulations that fit real world flow data and incorporate an appropriate uncertainty modeling.

#### • Control and optimization of turbulent flows

We are interested on active control and more precisely on closed-loop control. The main idea is to extract reliable image features to act on the flow. This approach is well known in the robot control community, it is called visual servoing. More generally, it is a technique to control a dynamic system from image features. We plan to apply this approach on flows involved in various domains such as environment, transport, microfluidic, industrial chemistry, pharmacy, food industry, agriculture, etc.

• Numerical models for geophysical flows simulation and analysis Numerical models are very useful for environmental applications. Several difficulties must be handled simultaneously, in a multidisciplinary context. For example, in geophysics, media are highly heterogeneous and only few data are available. Stochastic models are often necessary to describe unresolved physical processes. Computational domains are characterized by complex 3D geometries, requiring adapted space discretization. Equations modeling flow and transport are transient, requiring also adapted time discretization. Moreover, these equations can be coupled together or with other equations in a global nonlinear system. These large-scale models are very time and memory consuming. High performance computing is thus required to run these types of scientific simulations. Supercomputers and clusters are quite powerful, provided that the numerical models are written with a parallel paradigm.

# **3. Research Program**

# 3.1. Estimation of fluid characteristic features from images

The measurement of fluid representative features such as vector fields, potential functions or vorticity maps, enables physicists to have better understanding of experimental or geophysical fluid flows. Such measurements date back to one century and more but became an intensive subject of research since the emergence of correlation techniques [48] to track fluid movements in pairs of images of a particles laden fluid or by the way of clouds photometric pattern identification in meteorological images. In computer vision, the estimation of the projection of the apparent motion of a 3D scene onto the image plane, referred to in the literature as optical-flow, is an intensive subject of researches since the 80's and the seminal work of B. Horn and B. Schunk [61]. Unlike to dense optical flow estimators, the former approach provides techniques that supply only sparse velocity fields. These methods have demonstrated to be robust and to provide accurate measurements for flows seeded with particles. These restrictions and their inherent discrete local nature limit too much their use and prevent any evolutions of these techniques towards the devising of methods supplying physically consistent results and small scale velocity measurements. It does not authorize also the use of scalar images exploited in numerous situations to visualize flows (image showing the diffusion of a scalar such as dye, pollutant, light index refraction, flurocein,...). At the opposite, variational techniques enable in a well-established mathematical framework to estimate spatially continuous velocity fields, which should allow more properly to go towards the measurement of smaller motion scales. As these methods are defined through PDE's systems they allow quite naturally constraints to be included such as kinematic properties or dynamic laws governing the observed fluid flows. Besides, within this framework it is also much easier to define characteristic features estimation procedures on the basis of physically grounded data model that describes the relation linking the observed luminance function and some state variables of the observed flow. The Fluminance group has allowed a substantial progress in this direction with the design of dedicated dense estimation techniques to estimate dense fluid motion fields. See [7] for a detailed review. More recently problems related to scale measurement and uncertainty estimation have been investigated [55]. Dynamically consistent and highly robust techniques have been also proposed for the recovery of surface oceanic streams from satellite images [51]. Very recently parameter-free approaches relying on uncertainty concept has been devised [52]. This technique outperforms the state of the art.

# 3.2. Data assimilation and Tracking of characteristic fluid features

Real flows have an extent of complexity, even in carefully controlled experimental conditions, which prevents any set of sensors from providing enough information to describe them completely. Even with the highest levels of accuracy, space-time coverage and grid refinement, there will always remain at least a lack of resolution and some missing input about the actual boundary conditions. This is obviously true for the complex flows encountered in industrial and natural conditions, but remains also an obstacle even for standard academic flows thoroughly investigated in research conditions.

This unavoidable deficiency of the experimental techniques is nevertheless more and more compensated by numerical simulations. The parallel advances in sensors, acquisition, treatment and computer efficiency allow the mixing of experimental and simulated data produced at compatible scales in space and time. The inclusion of dynamical models as constraints of the data analysis process brings a guaranty of coherency based on fundamental equations known to correctly represent the dynamics of the flow (e.g. Navier Stokes equations) [11]. Conversely, the injection of experimental data into simulations ensures some fitting of the model with reality.

To enable data and models coupling to achieve its potential, some difficulties have to be tackled. It is in particular important to outline the fact that the coupling of dynamical models and image data are far from being straightforward. The first difficulty is related to the space of the physical model. As a matter of fact, physical models describe generally the phenomenon evolution in a 3D Cartesian space whereas images provides generally only 2D tomographic views or projections of the 3D space on the 2D image plane. Furthermore, these views are sometimes incomplete because of partial occlusions and the relations between the model state variables and the image intensity function are otherwise often intricate and only partially known. Besides, the dynamical model and the image data may be related to spatio-temporal scale spaces of very different natures which increases the complexity of an eventual multiscale coupling. As a consequence of these difficulties, it is necessary generally to define simpler dynamical models in order to assimilate image data. This redefinition can be done for instance on an uncertainty analysis basis, through physical considerations or by the way of data based empirical specifications. Such modeling comes to define inexact evolution laws and leads to the handling of stochastic dynamical models. The necessity to make use and define sound approximate models, the dimension of the state variables of interest and the complex relations linking the state variables and the intensity function, together with the potential applications described earlier constitute very stimulating issues for the design of efficient data-model coupling techniques based on image sequences.

On top of the problems mentioned above, the models exploited in assimilation techniques often suffer from some uncertainties on the parameters which define them. Hence, a new emerging field of research focuses on the characterization of the set of achievable solutions as a function of these uncertainties. This sort of characterization indeed turns out to be crucial for the relevant analysis of any simulation outputs or the correct interpretation of operational forecasting schemes. In this context, stochastic modeling play a crucial role to model and process uncertainty evolution along time. As a consequence, stochastic parameterization of flow dynamics has already been present in many contributions of the Fluminance group in the last years and will remain a cornerstone of the new methodologies investigated by the team in the domain of uncertainty characterization.

This wide theme of research problems is a central topic in our research group. As a matter of fact, such a coupling may rely on adequate instantaneous motion descriptors extracted with the help of the techniques studied in the first research axis of the FLUMINANCE group. In the same time, this coupling is also essential with respect to visual flow control studies explored in the third theme. The coupling between a dynamics and data, designated in the literature as a Data Assimilation issue, can be either conducted with optimal control techniques [62], [63] or through stochastic filtering approaches [56], [59]. These two frameworks have their own advantages and deficiencies. We rely indifferently on both approaches.

# 3.3. Optimization and control of fluid flows with visual servoing

Fluid flow control is a recent and active research domain. A significant part of the work carried out so far in that field has been dedicated to the control of the transition from laminarity to turbulence. Delaying, accelerating or modifying this transition is of great economical interest for industrial applications. For instance, it has been

shown that for an aircraft, a drag reduction can be obtained while enhancing the lift, leading consequently to limit fuel consumption. In contrast, in other application domains such as industrial chemistry, turbulence phenomena are encouraged to improve heat exchange, increase the mixing of chemical components and enhance chemical reactions. Similarly, in military and civilians applications where combustion is involved, the control of mixing by means of turbulence handling rouses a great interest, for example to limit infra-red signatures of fighter aircraft.

Flow control can be achieved in two different ways: passive or active control. Passive control provides a permanent action on a system. Most often it consists in optimizing shapes or in choosing suitable surfacing (see for example [54] where longitudinal riblets are used to reduce the drag caused by turbulence). The main problem with such an approach is that the control is, of course, inoperative when the system changes. Conversely, in active control the action is time varying and adapted to the current system's state. This approach requires an external energy to act on the system through actuators enabling a forcing on the flow through for instance blowing and suction actions [66], [58]. A closed-loop problem can be formulated as an optimal control issue where a control law minimizing an objective cost function (minimization of the drag, minimization of the actuators power, etc.) must be applied to the actuators [50]. Most of the works of the literature indeed comes back to open-loop control approaches [65], [60], [64] or to forcing approaches [57] with control laws acting without any feedback information on the flow actual state. In order for these methods to be operative, the model used to derive the control law must describe as accurately as possible the flow and all the eventual perturbations of the surrounding environment, which is very unlikely in real situations. In addition, as such approaches rely on a perfect model, a high computational costs is usually required. This inescapable pitfall has motivated a strong interest on model reduction. Their key advantage being that they can be specified empirically from the data and represent quite accurately, with only few modes, complex flows' dynamics. This motivates an important research axis in the Fluminance group.

# 3.4. Numerical models applied to hydrogeology and geophysics

The team is strongly involved in numerical models for hydrogeology and geophysics. There are many scientific challenges in the area of groundwater simulations. This interdisciplinary research is very fruitful with cross-fertilizing subjects.

In geophysics, a main concern is to solve inverse problems in order to fit the measured data with the model. Generally, this amounts to solve a linear or nonlinear least-squares problem.

Models of geophysics are in general coupled and multi-physics. For example, reactive transport couples advection-diffusion with chemistry. Here, the mathematical model is a set of nonlinear Partial Differential Algebraic Equations. At each timestep of an implicit scheme, a large nonlinear system of equations arise. The challenge is to solve efficiently and accurately these large nonlinear systems.

# 3.5. Numerical algorithms and high performance computing

Linear algebra is at the kernel of most scientific applications, in particular in physical or chemical engineering. The objectives are to analyze the complexity of these different methods, to accelerate convergence of iterative methods, to measure and improve the efficiency on parallel architectures, to define criteria of choice.

# 4. Application Domains

# 4.1. Introduction

By designing new approaches for the analysis of fluid-image sequences the FLUMINANCE group aims at contributing to several application domains of great interest for the community and in which the analysis of complex fluid flows plays a central role. The group focuses mainly on two broad application domains:

- Environmental sciences;
- Experimental fluid mechanics and industrial flows.

We detail hereafter these two application domains.

# 4.2. Environmental sciences

The first huge application domain concerns all the sciences that aim at observing the biosphere evolution such as meteorology, climatology or oceanography but also remote sensing study for the monitoring of meteorological events or human activities consequences. For all these domains image analysis is a practical and unique tool to *observe, detect, measure, characterize or analyze* the evolution of physical parameters over a large domain. The design of generic image processing techniques for all these domains might offer practical software tools to measure precisely the evolution of fluid flows for weather forecasting or climatology studies. It might also offer possibilities of close surveillance of human and natural activities in sensible areas such as forests, river edges, and valley in order to monitor pollution, floods or fire. The need in terms of local weather forecasting, risk prevention, or local climate change is becoming crucial for our tomorrow's life. At a more local scale, image sensors may also be of major utility to analyze precisely the effect of air curtains for safe packaging in agro-industrial.

# 4.3. Experimental fluid mechanics and industrial flows

In the domain of **experimental fluid mechanics**, the visualization of fluid flows plays a major role, especially for turbulence study since high frequency imaging has been made currently available. Together with analysis of turbulence at different scales, one of the major goals pursued at the moment by many scientists and engineers consists in studying the ability to manipulate a flow to induce a desired change. This is of huge technological importance to enhance or inhibit mixing in shear flows, improve energetic efficiency or control the physical effects of strain and stresses. This is for instance of particular interest for:

- military applications, for example to limit the infra-red signatures of fighter aircraft;
- aeronautics and transportation, to limit fuel consumption by controlling drag and lift effects of turbulence and boundary layer behavior;
- industrial applications, for example to monitor flowing, melting, mixing or swelling of processed materials, or preserve manufactured products from contamination by airborne pollutants, or in industrial chemistry to increase chemical reactions by acting on turbulence phenomena.

# 5. Highlights of the Year

# 5.1. Highlights of the Year

# 5.1.1. Awards

First prize of the second Mathematics of Planet Earth international competition. Module "Simulating the melting of ice caps", authors M. Nodet and J. Erhel. BEST PAPERS AWARDS :

DESTTAFERS AWARDS.

[] Simulating the melting of ice caps.

# 6. New Software and Platforms

# 6.1. 2DLayeredMotion

Estimation of 2D independent mesoscale layered atmospheric motion fields

FUNCTIONAL DESCRIPTION: This software enables to estimate a stack of 2D horizontal wind fields corresponding to a mesoscale dynamics of atmospheric pressure layers. This estimator is formulated as the minimization of a global energy function. It relies on a vertical decomposition of the atmosphere into pressure layers. This estimator uses pressure data and classification clouds maps and top of clouds pressure maps (or infra-red images). All these images are routinely supplied by the EUMETSAT consortium which handles the Meteosat and MSG satellite data distribution. The energy function relies on a data model built from the integration of the mass conservation on each layer. The estimator also includes a simplified and filtered shallow water dynamical model as temporal smoother and second-order div-curl spatial regularizer. The estimator may also incorporate correlation-based vector fields as additional observations. These correlation vectors are also routinely provided by the Eumetsat consortium.

- Participant: Étienne Mémin
- Contact: Étienne Mémin
- URL: http://fluid.irisa.fr/index.html

# 6.2. 3DLayeredMotion

#### Estimation of 3D interconnected layered atmospheric motion fields

FUNCTIONAL DESCRIPTION: This software extends the previous 2D version. It allows (for the first time to our knowledge) the recovery of 3D wind fields from satellite image sequences. As with the previous techniques, the atmosphere is decomposed into a stack of pressure layers. The estimation relies also on pressure data and classification clouds maps and top of clouds pressure maps. In order to recover the 3D missing velocity information, physical knowledge on 3D mass exchanges between layers has been introduced in the data model. The corresponding data model appears to be a generalization of the previous data model constructed from a vertical integration of the continuity equation.

- Contact: Étienne Mémin
- URL: http://fluid.irisa.fr

# 6.3. DenseMotion

#### Estimation of 2D dense motion fields

FUNCTIONAL DESCRIPTION: This code allows the computation from two consecutive images of a dense motion field. The estimator is expressed as a global energy function minimization. The code enables the choice of different data models and different regularization functionals depending on the targeted application. Generic motion estimators for video sequences or fluid flows dedicated estimators can be set up. This software allows in addition the users to specify additional correlation based matching measurements. It enables also the inclusion of a temporal smoothing prior relying on a velocity vorticity formulation of the Navier-Stoke equation for Fluid motion analysis applications.

- Participant: Étienne Mémin
- Contact: Étienne Mémin
- URL: http://fluid.irisa.fr/index.html

# 6.4. Low-Order-Motion

Estimation of low order representation of fluid motion

FUNCTIONAL DESCRIPTION: This code enables the estimation of a low order representation of a fluid motion field from two consecutive images. The fluid motion representation is obtained using a discretization of the vorticity and divergence maps through regularized Dirac measure. The irrotational and solenoidal components of the motion fields are expressed as linear combinations of basis functions obtained through the Biot-Savart law. The coefficient values and the basis function parameters are formalized as the minimizer of a functional relying on an intensity variation model obtained from an integrated version of the mass conservation principle of fluid mechanics.

- Participants: Anne Cuzol and Étienne Mémin
- Contact: Étienne Mémin
- URL: http://fluid.irisa.fr

# 6.5. TYPHOON

- Participants: Christopher Mauzey, Étienne Mémin and Pierre Dérian
- Partner: CSU Chico
- Contact: Étienne Mémin
- URL: http://phys.csuchico.edu/lidar/typhoon/

# 6.6. H2OLab

KEYWORDS: Energy - Contamination - Groundwater - Hydrogeology - Heterogeneity - Uncertainly - Multiscale - Simulation

SCIENTIFIC DESCRIPTION: The software platform contains a database which is interfaced through the web portal H2OWeb. It contains also software modules which can be used through the interface H2OGuilde. The platform H2OLab is an essential tool for the dissemination of scientific results. Currently, software and database are shared by the partners of the h2mno4 project.

FUNCTIONAL DESCRIPTION: The software platform H2OLab is devoted to stochastic simulations of groundwater flow and contaminant transport in highly heterogeneous porous and fractured geological media.

-Modeling and numerical simulation of aquifers -Porous and fractured heterogeneous media -Flow with mixed finite elements -Solute transport with a Lagrangian method -Stochastic modeling for data uncertainty.

- Participants: Géraldine Pichot, Grégoire Lecourt, Jean-Raynald De Dreuzy and Jocelyne Erhel
- Partners: Université de Rennes 1 CNRS Université de Lyon Université de Poitiers
- Contact: Jocelyne Erhel
- URL: http://h2olab.inria.fr/

# 6.7. PALMTREE

KEYWORD: Monte-Clarlo

FUNCTIONAL DESCRIPTION: We present an easy-to-use package for the parallelization of Lagrangian methods for partial differential equations. In addition to the reduction of computation time, the code aims at satisfying three properties:

simplicity: the user just has to add the algorithm governing the behaviour of the particles. portability: the possibility to use the package with any compiler and OS. action-replay: the ability of the package to replay a selected batch of particles.

The last property allows the user to replay and capture the whole sample path for selected particles of a batch. This feature is very useful for debugging and catching some relevant information.

- Authors: Lionel Lenôtre, Géraldine Pichot, Lionel Lenôtre and Lionel Lenôtre
- Contact: Géraldine Pichot

# 6.8. GRT3D

KEYWORDS: Geochemistry - Dispersion - Scientific calculation - Simulation - Advection SCIENTIFIC DESCRIPTION: Participants : Édouard Canot, Jocelyne Erhel [correspondant].

Version: version 2.0, April 2014

APP: registered

Programming language: C

Abstract: Reactive transport modeling has become an essential tool for understanding complex environmental problems. It is an important issue for MoMaS and C2S@EXA partners (see sections 8.2.5, 8.2.3), in particular Andra. We have developed a method coupling transport and chemistry, based on a method of lines such that spatial discretization leads to a semi-discrete system of algebraic differential equations (DAE system). The main advantage is to use a complex DAE solver, which controls simultaneously the timestep and the convergence of Newton algorithm. The approach SIA uses a fixed-point method to solve the nonlinear system at each timestep, whereas the approach SNIA uses an explicit scheme.

The software suite GRT3D has four executable modules:

SIA1D: Sequential Iterative Approach for 1D domains,

GDAE1D: Global DAE approach for 1D domains,

SNIA3D: Sequential Non Iterative Approach for 1D, 2D or 3D domains.

GDAE3D: Global DAE approach for 1D, 2D or 3D domains. This module has three variants: the original one with logarithms, an optimized one still with logarithms, an optimized one which does not use logarithms.

Current work: extension of the chemistry module and parallelization.

FUNCTIONAL DESCRIPTION: Reactive transport modeling has become an essential tool for understanding complex environmental problems. It is an important issue for MoMaS and C2S@EXA partners, in particular Andra. We have developed a method coupling transport and chemistry, based on a method of lines such that spatial discretization leads to a semi-discrete system of algebraic differential equations (DAE system). The main advantage is to use a complex DAE solver, which controls simultaneously the timestep and the convergence of Newton algorithm. The approach SIA uses a fixed-point method to solve the nonlinear system at each timestep, whereas the approach SNIA uses an explicit scheme.

The software suite GRT3D has four executable modules:

SIA1D: Sequential Iterative Approach for 1D domains,

GDAE1D: Global DAE approach for 1D domains,

SNIA3D: Sequential Non Iterative Approach for 1D, 2D or 3D domains.

GDAE3D: Global DAE approach for 1D, 2D or 3D domains. This module has three variants: the original one with logarithms, an optimized one still with logarithms, an optimized one which does not use logarithms.

- Participants: Caroline De Dieuleveult, Édouard Canot, Jocelyne Erhel, Nadir Soualem and Souhila Sabit
- Partner: ANDRA
- Contact: Jocelyne Erhel

# 7. New Results

# 7.1. Fluid motion estimation

### 7.1.1. Stochastic uncertainty models for motion estimation

Participants: Shengze Cai, Etienne Mémin, Musaab Khalid Osman Mohammed.

The objective consists here in relying on a stochastic transport formulation to propose a luminance conservation assumption dedicated to the measurement of large-scale fluid flows velocity. This formulation, relying on the modeling under location uncertainty principle developed in the team, has the great advantage to incorporate from the beginning an uncertainty on the unresolved (turbulent) motion measurement. This uncertainty modeled as a possibly inhomogeneous random field uncorrelated in time can be estimated jointly to the motion estimates. Such a formulation, besides providing estimates of the velocity field and of its associated uncertainties, allows us to naturally define a linear multiresolution scale-space framework. It provides also a reinterpretation, in terms of uncertainty, of classical regularization functionals proposed in the context of motion estimation. Nevertheless, at variance to classical motion estimation methods, this approach enables to estimate the so-called regularization parameter, which is in our framework related to the variance of the unresolved component of motion component. The resulting parameter-free estimator has shown to outperform state-of-the-art results of the literature [14]. This kind of method is applied on turbulent flows and in the context of river hydrologic applications through a collaboration with the Irstea Lyon research group (HHLY). A method for the 3D reconstruction of the river plane has been also proposed in this context. This study is performed within the PhD thesis of Musaab Mohammed.

### 7.1.2. Surface Currents estimation from Shore-Based Videos

### Participant: Pierre Derian.

A wavelet based motion estimator has been extended for the recovery of instantaneous fields of surface current from shore-based and unmanned aerial vehicle videos. This study published in [16] and [34] demonstrated clearly the high potential of this method in the nearshore, where the rapid development of webcams and drones offers a large amount of applications for swimming and surfing safety, engineering and naval security and research purpose, by providing quantitative information. This work has been conducted within a collaboration with the Legos laboratory.

# 7.1.3. Development of an image-based measurement method for large-scale characterization of indoor airflows

Participants: Dominique Heitz, Etienne Mémin, Romain Schuster.

The goal is to design a new image-based flow measurement method for large-scale industrial applications. From this point of view, providing *in situ* measurement technique requires: (i) the development of precise models relating the large-scale flow observations to the velocity; (ii) appropriate large-scale regularization strategies; and (iii) adapted seeding and lighting systems, like Hellium Filled Soap Bubles (HFSB) and led ramp lighting. This work conducted within the PhD of Romain Schuster in collaboration with the compagny ITGA has started in february 2016. The first step has been to evaluate the performances of a stochastic uncertainty motion estimator when using large scale scalar images, like those obtained when seeding a flow with smoke. The PIV characterization of flows on large fields of view requires an adaptation of the motion estimation method from image sequences. The backward shift of the camera coupled to a dense scalar seeding involves a large scale observation of the flow, thereby producing uncertainty about the observed phenomena. By introducing a stochastic term related to this uncertainty into the observation term, we obtained a significant improvement of the estimated velocity field accuracy [41].

### 7.1.4. 3D flows reconstruction from image data

Participants: Dominique Heitz, Cédric Herzet.

Our work focuses on the design of new tools for the estimation of 3D turbulent flow motion in the experimental setup of Tomo-PIV. This task includes both the study of physically-sound models on the observations and the fluid motion, and the design of low-complexity and accurate estimation algorithms.

This year, we continued our investigation on the problem of efficient volume reconstruction. During the last years, we have focussed our attention on several families of convex optimization algorithms allowing to accelerate the standard procedures encountered in the Tomo-PIV literature while accounting for the non-negativity and the sparsity of the sought solutions. So far, the assessment of the proposed algorithms were exclusively done on synthetic data. This year, we started the process of validating the proposed procedures on real experimental data.

We started through a collaboration with Irstea to study ensemble assimilation methods for the fast reconstruction of 3D tomo-PIV motion field. The approach relies on a simplified dynamics of the flow and is a generalization of one of the popular emerging flow reconstruction technique of the PIV community referred to as "Shake the box". The study will be developed within an Irstea post-doctoral funding.

### 7.1.5. Sparse-representation algorithms

#### Participant: Cédric Herzet.

The paradigm of sparse representations is a central concept in many domains of signal processing. In particular, in the field of fluid motion estimation, sparse representation appears to be potentially useful at several levels: (i) it provides a relevant model for the characterization of the velocity field in some scenarios; (ii) it plays a crucial role in the recovery of volumes of particles in the 3D Tomo-PIV problem.

Unfortunately, the standard sparse representation problem is known to be NP hard. Therefore, heuristic procedures have to be devised to access to the solution of this problem. This year, we continued our investigations on "screening" methodologies, that is procedures allowing for the rapid identification of (some of) the zeros of the sought sparse vector. More specifically, we designed low-complexity procedures enabling to screen groups of atoms by only performing one single test. This work has been submitted to the IEEE international conference on acoustic, speech and signal processing (ICASSP).

# 7.2. Tracking, Data assimilation and model-data coupling

# 7.2.1. Optimal control techniques for the coupling of large scale dynamical systems and image data

Participants: Mohamed Yacine Ben Ali, Pranav Chandramouli, Dominique Heitz, Etienne Mémin.

In this axis of work we are exploring the use of optimal control techniques for the coupling of Large Eddies Simulation (LES) techniques and 2D image data. The objective is to reconstruct a 3D flow from a set of simultaneous time resolved 2D image sequences visualizing the flow on a set of 2D plans enlightened with laser sheets. This approach is experimented on shear layer flows and on wake flows generated on the wind tunnel of Irstea Rennes. Within this study we aim to explore techniques to enrich large-scale dynamical models by the introduction of uncertainty terms or through the definition of subgrid models from the image data. This research theme is related to the issue of turbulence characterization from image sequences. Instead of predefined turbulence models, we aim here at tuning from the data the value of coefficients involved in traditional LES subgrid models. A 4DVar assimilation technique based on the numerical code Incompact3D has been implemented for that purpose to control the inlet and initial conditions in order to reconstruct a turbulent wake flow behind an unknown obstacle. We are actually extending this first data assimilation technique to control the subgrid parameters. This study is performed in collaboration with Sylvain Laizet (Imperial College). In another axis of research, in collaboration with the CSTB Nantes centre and within the PhD of Yacine Ben Ali we will explore the definition of efficient data assimilation schemes for wind engineering. The goal will be here to couple Reynolds average model to pressure data at the surface of buildings. The final purpose will consist in proposing improved data-driven simulation models for architects.

# 7.2.2. Ensemble variational data assimilation of large-scale dynamics with uncertainty Participant: Etienne Mémin.

This study is focused on the coupling of a large scale representation of the flow dynamics built from the location uncertainty principle with image data of finer resolution. The velocity field at large scales is described as a regular smooth component whereas the complement component is a highly oscillating random velocity field defined on the image grid but living at all the scales. Following this route we have assessed the performance of an ensemble variational assimilation technique with direct image data observation. Good results have been obtained for simulation under location uncertainty of 1D and 2D shallow water models [26]. This open the way to the definition of efficient data assimilation schemes for the coupling of high resolution data with large scale dynamical system.

### 7.2.3. Reduced-order models for flows representation from image data

Participants: Mamadou Diallo, Dominique Heitz, Cédric Herzet, Etienne Mémin, Valentin Resseguier.

During the PhD thesis of Valentin Resseguier we proposed a new decomposition of the fluid velocity in terms of a large-scale continuous component with respect to time and a small-scale non continuous random component. Within this general framework, an uncertainty based representation of the Reynolds transport theorem and Navier-Stokes equations can be derived, based on physical conservation laws. This physically relevant stochastic model has been applied in the context of POD-Galerkin methods. This uncertainty modeling methodology provides a theoretically grounded technique to define an appropriate subgrid tensor as well drift correction terms. The pertinence of this stochastic reduced order model has been successfully assessed on several wake flows at different Reynold number. It has been shown to be much more stable than the usual reduced order model construction techniques. Beyond the definition of a stable reduced order model, the modeling under location uncertainty paradigm offers a unique way to analyse from the data of a turbulent flow the action of the small-scale velocity components on the large-scale flow [25]. Regions of prominent turbulent kinetic energy, direction of preferential diffusion, as well as the small-scale induced drift can be identified and analyzed to decipher key players involved in the flow. This study has been published in the journal of fluid mechanics. Note that these reduced order models can be extended to a full system of stochastic differential equation driving all the temporal modes of the reduced system (and not only the small-scale modes). This full stochastic system has been evaluated on wake flow at moderate Reynolds number. For this flow the system has shown to provide very good uncertainty quantification properties as well meaningful physical behavior with respect to the simulation of the neutral modes of the dynamics. This study described in the PhD of Valentin Resseguier will be soon submitted to a journal paper.

On the other hand, in the following of several approaches proposed by the team [49], [53], we continued our investigations on the estimation of deterministic reduced order model from partial observations. In this line of research, we proposed a Bayesian framework for the construction of reduced-order models from image data. Our framework combines observation and prior information on the system to reduce the model and takes into account the uncertainties on the parameters of the model. The proposed approach reduces to some well-known model-reduction techniques for complete observations (i.e., the observation operator can be inverted). A theoretical analysis of our methodology has been investigated in a simplified context (namely, the observations are supposed to be noiseless linear combinations of the state of the system). This result provides worst-case guarantees on the reconstruction performance which can be achieved by a reduced model built from the data. These contributions have led to publications in a journal [18] and a conference [33].

# 7.3. Analysis and modeling of turbulent flows and geophysical flows

### 7.3.1. Geophysical flows modeling under location uncertainty

Participants: Pierre Derian, Long Li, Etienne Mémin, Valentin Resseguier.

In this research axis we have devised a principle to derive representation of flow dynamics under uncertainty. Such an uncertainty is formalized through the introduction of a random term that enables taking into account large-scale approximations or truncation effects performed within the dynamics analytical constitution steps. This includes for instance the modeling of unresolved scales interaction in large eddies simulation (LES) or in Reynolds average numerical simulation (RANS), but also partially known forcing. Rigorously derived from a stochastic version of the Reynolds transport theorem [9], this framework, referred to as modeling under location uncertainty, encompasses several meaningful mechanisms for turbulence modeling. It indeed introduces without any supplementary assumption the following pertinent mechanisms for turbulence modeling: (i) a dissipative operator related to the mixing effect of the large-scale components by the small-scale velocity; (ii) a multiplicative noise representing small-scale energy backscattering; and (iii) a modified advection term related to the so-called *turbophoresis* phenomena, attached to the migration of inertial particles in regions of lower turbulent diffusivity.

In a series of papers we have shown how such modeling can be applied to provide stochastic representations of a variety of classical geophysical flows dynamics [24], [23], [22]. Numerical simulations and uncertainty quantification have been performed on Quasi Geostophic approximation (QG) of oceanic models. It has been shown that such models lead to remarkable estimation of the unresolved errors at variance to classical eddy viscosity based models. The noise brings also an additional degree of freedom in the modeling step and pertinent diagnostic relations and variations of the model can be obtained with different scaling assumptions of the turbulent kinetic energy (i.e. of the noise amplitude). The performances of such systems have been assessed also on an original stochastic representation of the Lorenz 63 derived from the modeling under location uncertainty [15]. In this study it has been shown that the stochastic version enabled to explore in a much faster way the region of the deterministic attractor. This effort has been undertaken within a fruitful collaboration with Bertrand Chapron (LOPS/IFREMER). In the PhD of Long Li, starting this year, we will continue this effort. The goal will be to propose relevant techniques to define or calibrate the noise term from data. In that prospect, we intend to explore statistical learning techniques.

### 7.3.2. Large eddies simulation models under location uncertainty

Participants: Mohamed Yacine Ben Ali, Pranav Chandramouli, Dominique Heitz, Etienne Mémin.

The models under location uncertainty recently introduced by Mémin (2014) [9] provide a new outlook on LES modeling for turbulence studies. These models are derived from a stochastic transport principle. The associated stochastic conservation equations are similar to the filtered Navier- Stokes equation wherein we observe a sub-grid scale dissipation term. However, in the stochastic version, an extra term appears, termed as "velocity bias", which can be treated as a biasing/modification of the large-scale advection by the small scales. This velocity bias, introduced artificially in the literature, appears here automatically through a decorrelation assumption of the small scales at the resolved scale. All sub-grid contributions for the stochastic models are defined by the small-scale velocity auto-correlation tensor. This large scale modeling has been assed and compared to several classical large-scale models on several flows, namely a flow over a circular cylinder at Re  $\sim 3900$  [32], a smooth channel flow at Re(tau)  $\sim 395$  [31] and Taylor-Green vortex flows at Reynolds 1600, 3000 and 5000 [20]. For all these flows the modeling under uncertainty has provided better results than classical large eddies simulation models. Within the PhD of Yacine Ben Ali we will explore with the CSTB Nantes centre the application of such models for the definition of Reynolds average simulation (RANS) models for wind engineering applications.

# 7.3.3. Singular and regular solutions to the Navier-Stokes equations (NSE) and relative turbulent models

Participants: Roger Lewandowski, Etienne Mémin, Benoit Pinier.

The common thread of this work is the problem set by J. Leray in 1934 : does a regular solution of the Navier-Stokes equations (NSE) with a smooth initial data develop a singularity in finite time, what is the precise structure of a global weak solution to the Navier-Stokes equations, and are we able to prove any uniqueness result of such a solution. This is a very hard problem for which there is for the moment no answer. Nevertheless, this question leads us to reconsider the theory of Leray for the study of the Navier-Stokes equations in the whole space with an additional eddy viscosity term that models the Reynolds stress in the context of large-scale flow modelling. It appears that Leray's theory cannot be generalized turnkey for this problem, so that things must be reconsidered from the beginning. This problem is approached by a regularization process using mollifiers, and particular attention must be paid to the eddy viscosity term. For this regularized problem and when the eddy viscosity has enough regularity, we have been able to prove the existence of a global unique solution that is of class  $C^{\infty}$  in time and space and that satisfies the energy balance. Moreover, when the eddy viscosity is of compact support in space, uniformly in time, we recently shown that this solution converges to a turbulent solution to the corresponding Navier-Stokes equations when, the regularizing parameter goes to 0. These results are described in a paper that has been submitted to the journal Archive for Rational Mechanics and Analysis (ARMA).

In the same direction, we also finalized a paper in collaboration with L. Berselli (Univ. Pisa, Italy) about the well known Bardina's turbulent model. In this problem, we consider the Helmholtz filter usually used within the framework of Large Eddy Simulation. We carry out a similar analysis, by showing in particular that no singularity occurs for Bardina's model.

Another study in collaboration with B. Pinier, P. Chandramouli and E. Memin has been undertaken. This work takes place within the context of the PhD work of B. Pinier. We considered the standard turbulent models involving the Navier-Stokes equations with an eddy viscosity that depends on the Turbulent Kinetic Energy (TKE), coupled with a supplementary equation for the TKE. The problem holds in a 3D bounded domain, with the Manning law at the boundary for the velocity. We have modeled a flux condition at the boundary for the TKE. We prove that with these boundary conditions, the resulting problem has a distributional solution. Then a serie of numerical tests has been performed in a parallelepiped with a non trivial bottom, showing the accuracy of the model in comparison with a direct numerical simulation of the Navier-Stokes equations. This work will be submitted to a journal.

### 7.3.4. Turbulence similarity theory for the modeling of Ocean Atmosphere interface

Participants: Roger Lewandowski, Etienne Mémin, Benoit Pinier.

The Ocean Atmosphere interface plays a major role in climate dynamics. This interaction takes place in a thin turbulent layer. To date no sastifying universal models for the coupling of atmospheric and oceanic models exists. In practice this coupling is realized through empirically derived interaction bulks. In this study, corresponding to the PhD thesis of Benoit Pinier, we aim at exploring similarity theory to identify universal mean profile of velocity and temperature within the mixture layer. The goal of this work consists in exhibiting eddy viscosity models within the primitive equations. We will also explore the links between those eddy viscocity models and the subgrid tensor derived from the uncertainty framework studied in the Fluminance group. In that prospect, we have studied the impact of the introduction of a random modeling of the friction velocity on the classical wall law expression. This model derived within the modeling under location uncertainty principle enabled us to propose an improved model of the velocity profile with a clear formulation in particular in the buffer turbulent area between the viscous zone and the turbulent region. Preliminary results on chanel flows are very promising. We are actually assessing this model on turbulent boundary layer flow at high Reynold.

# 7.3.5. Hot-wire anemometry at low velocities

Participant: Dominique Heitz.

A new dynamical calibration technique has been developed for hot-wire probes. The technique permits, in a short time range, the combined calibration of velocity, temperature and direction calibration of single and multiple hot-wire probes. The calibration and measurements uncertainties were modeled, simulated and controlled, in order to reduce their estimated values. Based on a market study the french patent application has been extended this year to a Patent Cooperation Treaty (PCT) application.

### 7.3.6. Numerical and experimental image and flow database

Participants: Pranav Chandramouli, Dominique Heitz.

The goal was to design a database for the evaluation of the different techniques developed in the Fluminance group. The first challenge was to enlarge a database mainly based on two-dimensional flows, with threedimensional turbulent flows. Synthetic image sequences based on homogeneous isotropic turbulence and on circular cylinder wake have been provided. These images have been completed with time resolved Particle Image Velocimetry measurements in wake and mixing layers flows. This database provides different realistic conditions to analyse the performance of the methods: time steps between images, level of noise, Reynolds number, large-scale images. The second challenge was to carried out orthogonal dual plane time resolved stereoscopic PIV measurements in turbulent flows. The diagnostic employed two orthogonal and synchronized stereoscopic PIV measurements to provide the three velocity components in planes perpendicular and parallel to the streamwise flow direction. These temporally resolved planar slices observations will be used in 4DVar assimilation technique, integrating Direct Numerical Simulation (DNS) and Large Eddies Simulation (LES), to reconstruct three-dimensional turbulent flows. This reconstruction will be conducted within the PhD of Pranav Chandramouli. The third challenge was to carried out a time resolved tomoPIV experiments in a turbulent wake flow. Then this data will be used to assess the performances of the 4DVar assimilation technique developed in the context of Pranav Chandramouli's PhD to reconstruct three-dimensional turbulent flows.

# 7.4. Visual servoing approach for fluid flow control

### 7.4.1. Closed-loop control of a spatially developing shear layer

Participants: Christophe Collewet, Johan Carlier.

This study aims at controlling one of the prototypical flow configurations encountered in fluid mechanics: the spatially developing turbulent shear layer occuring between two parallel incident streams with different velocities. Our goal is to maintain the shear-layer in a desired state and thus to reject upstream perturbations. In our conference IFAC paper (https://hal.inria.fr/hal-01514361) we focused on perturbations belonging to the same space that the actuators, concretely that means that we were only able to face perturbations of the actuator itself, like failures of the actuator. This year we enlarged this result to purely exogenous perturbations. An optimal control law has been derived to minimize the influence of the perturbation on the flow. To do that, an on-line estimation of the perturbation has been used. This work will be submitted to the upcoming IEEE Conference on Decision and Control. We have also generalized the works initiated during the post-doctoral stay of Tudor-Bogdan Airimitoaie (https://hal.archives-ouvertes.fr/hal-01101089) concerning the benefits of increasing the controlled degrees of freedom in the particular case of the heat equation. This work has been validated on the shear flow.

# 7.5. Coupled models in hydrogeology

### 7.5.1. Coupling of subsurface and seepage flows

Participants: Jocelyne Erhel, Jean-Raynald de Dreuzy.

Hillslope response to precipitations is characterized by sharp transitions from purely subsurface flow dynamics to simultaneous surface and subsurface flows. Locally, the transition between these two regimes is triggered by soil saturation. Here we develop an integrative approach to simultaneously solve the sub- surface flow, locate the potential fully saturated areas and deduce the generated saturation excess over- land flow. This approach combines the different dynamics and transitions in a single partition formulation using discontinuous functions. We propose to regularize the system of partial differential equations and to use classic spatial and temporal discretization schemes. We illustrate our methodology on the 1D hillslope storage Boussinesq equations. We first validate the numerical scheme on previous numerical experiments without saturation excess overland flow. Then we apply our model to a test case with dynamic transitions from purely subsurface flow dynamics to simultaneous surface and subsurface flows. Our results show that discretization respects mass balance both locally and globally, converges when the mesh or time step are refined. Moreover the regularization parameter can be taken small enough to ensure accuracy without suffering of numerical artefacts. Applied to some hundreds of realistic hillslope cases taken from Western side of France (Brittany), the developed method appears to be robust and efficient. This study performed within the H2MNO4 ANR project has been published in the journal Advances in Water Ressources [21].

### 7.5.2. Characterizations of Solutions in Geochemistry

#### Participant: Jocelyne Erhel.

We study the properties of a geochemical model involving aqueous and precipitation-dissolution reactions at a local equilibrium. By reformulating the model as an equivalent optimization problem, we prove existence and uniqueness of a solution. It is classical in thermodynamic to compute diagrams representing the phases of the system. We introduce here the new precipitation diagram that describes the mineral speciation in function of the parameters of the system. Using the polynomial structure of the problem, we provide characterizations and an algorithm to compute the precipitation diagram. Numerical computations on some examples illustrate this approach. This work, is part of the H2MNO4 initiative. It has been recently submitted to a journal for publication [45].

### 7.5.3. Reactive transport in fractured-porous media

Participants: Yvan Crenner, Jean-Raynald de Dreuzy, Jocelyne Erhel.

Fractures must be carefully considered for the geological disposal of radioactive wastes. They critically enhance diffusivity, speed up solute transport, extend mixing fronts, and in turn, modify the physico-chemical conditions of reactivity in the Excavation Damaged Zone (EDZ) of the galleries. On the other hand, the pyrite oxidation could be considered like the main reaction due to the diffusion of oxygen through the gallery. Moreover, we assume that this reaction is complete in these geological conditions. First, we propose a numerical explicit reactive transport model in a fractured medium for an overall reaction. Afterwards, we present simulations outputs of the pyrite-oxygen reaction in EDZ zone. This study supported by ANDRA has been published in a conference [27].

## 7.5.4. Reactive transfers for multi-phasic flows

Participants: Jocelyne Erhel, Bastien Hamlat.

This study focuses on the mathematical modeling of reactive transfers for multi-phasic flows in porous medium. This study supported by IFPEN has been presented in a conference paper [37].

# 7.6. Linear solvers

### 7.6.1. Variable s-step GMRES

Participants: Jocelyne Erhel, David Imberti.

Sparse linear systems arise in computational science and engineering. The goal is to reduce the memory requirements and the computational cost, by means of high performance computing algorithms. We introduce a new variation on s-step GMRES in order to improve its stability, reduce the number of iterations necessary to ensure convergence, and thereby improve parallel performance. In doing so, we develop a new block variant that allows us to express the stability difficulties in s-step GMRES more fully. This work supported by the EoCoE grant has been published in a conference proceeding [38] and in the journal [28].

### 7.6.2. Krylov method applied to reactive transport

#### Participant: Jocelyne Erhel.

Reactive transport models couple advection dispersion equations with chemistry equations. If the reactions are at thermodynamic equilibrium, then the system is a set of partial differential and algebraic equations. After space and implicit time discretizations, a nonlinear system of equations must be solved at each time step. The Jacobian matrix of the nonlinear system can be written with a Kronecker product coupling transport and chemistry. Krylov methods are well-suited to solve such linear systems because the matrix vector product can be done efficiently. The main challenge is to design a preconditioning matrix. We propose here to use the special structure of the matrix. Preliminary experiments show that Krylov methods are much more efficient than a direct method which does not use the coupled structure. This work supported by ANDRA has been published at the occasion of an invited conference [28].

# 8. Bilateral Contracts and Grants with Industry

# 8.1. Bilateral Contracts with Industry

# 8.1.1. Contract CERSAT/IFREMER

Participants: Etienne Mémin, Valentin Resseguier.

*duration 39 months.* This partnership between Inria and Ifremer funded the PhD of Valentin Resseguier, which aimed to study image based data assimilation strategies for oceanic models incorporating random uncertainty terms. The goal targeted will consist in deriving appropriate stochastic version of oceanic model and on top of them to devise estimation procedures from noisy data to calibrate the associated subgrid models.
# 8.1.2. Contract inter Carno IFREMER Inria

Participants: Etienne Mémin, Thibaut Tronchin.

*duration 36 months*. This contract was centred on the elaboration of an image-based tools for the analysis of the hydraulic load of an immersed body. This project tooke place within an inter Carnot cooperation between Ifremer and Inria.

# 8.1.3. Contract ITGA

Participants: Dominique Heitz, Etienne Mémin.

*duration 36 months*. This partnership between Inria, Irstea and ITGA funds the PhD of Romain Schuster. The goal of this PhD is to design new image-based flow measurement methods for the study of industrial fluid flows. Those techniques will be used in particular to calibrate industrial fume hood.

# 8.1.4. Contract CSTB

Participants: Mohamed Yacine Ben Ali, Dominique Heitz, Etienne Mémin.

*duration 36 months.* This partnership between Inria, Irstea and CSTB funds the PhD of Yacine Ben Ali. This PhD aims to design new data assimilation scheme for Reynolds Average Simulation (RANS) of flows involved in wind engineering and buildings construction. The goal pursued here consists to couple RANS models and surface pressure data in order to define data driven models with accurate turbulent parameterization.

# 8.1.5. ANDRA project

Participants: Yvan Crenner, Benjamin Delfino, Jean-Raynald de Dreuzy, Jocelyne Erhel.

Contract with ANDRA (National Agency for Nuclear Waste)

Duration: three years from November 2015.

Title: reactive transport in fractured porous media

Coordination: Jocelyne Erhel.

Partners: Geosciences Rennes.

Abstract: Even in small numbers, fractures must be carefully considered for the geological disposal of radioactive waste. They critically enhance diffusivity, speed up solute transport, extend mixing fronts and, in turn, modify the physicochemical conditions of reactivity around possible storage sites. Numerous studies in various fields have shown that fractures cannot be simply integrated within an equivalent porous medium with a simple enhancement of its petro-physical properties (porosity and permeability). We propose a combined numerical and experimental approach to determine the influence on reactivity of typical fracture patterns found in some radioactive waste applications.

# 8.1.6. IFPEN project

Participants: Bastien Hamlat, Jocelyne Erhel.

Contract with IFPEN (Institut Français du Pétrole et Energies Nouvelles)

Duration: three years from October 2016.

Title: Fully implicit Formulations for the Simulation of Multiphase Flow and Reactive Transport Coordination: Jocelyne Erhel.

Abstract: Modeling multiphase flow in porous media coupled with fluid-rock chemical reactions is essential in order to understand the origin of sub-surface natural resources and optimize their extraction. This project aims to determine optimal strategies to solve the coupled transport and chemical reaction equations describing the physical processes at work in reactive multiphase flow in porous media. Three different formulations show great potential to accurately solve these equations. Two are fully implicit ("Reactive Coats" and "Semi-smooth Newton)" and one is an operator splitting approach. These formulations are still incomplete at the moment. The work will focus on extending the existing formulations to more complex physical phenomena, study their stability, convergence and theoretical equivalence. Another objective is to provide practical solutions to efficiently solve the resulting non-linear systems.

# 9. Partnerships and Cooperations

# 9.1. National Initiatives

# 9.1.1. Comins'lab: SEACS : Stochastic modEl-dAta-Coupled representationS for the analysis, simulation and reconstruction of upper ocean dynamics

Participants: Pierre Derian, Cédric Herzet, Etienne Mémin.

*duration 48 months.* The SEACS project whose acronym stands for: "Stochastic modEl-dAta-Coupled representationS for the analysis, simulation and reconstruction of upper ocean dynamics" is a Joint Research Initiative between the three Britanny clusters of excellence of the "Laboratoires d'Excellence" program: Cominlabs, Lebesgue and LabexMer centered on numerical sciences, mathematics and oceanography respectively. Within this project we aim at studying the potential of large-scale oceanic dynamics modeling under uncertainty for ensemble forecasting and satellite image data assimilation.

# 9.1.2. ANR JCJC GERONIMO : Advanced GEophysical Reduced-Order Model construction from IMage Observations

Participants: Mamadou Diallo, Cédric Herzet.

*duration 48 months.* The GERONIMO project which started in March 2014 aims at devising new efficient and effective techniques for the design of geophysical reduced-order models from image data. The project both arises from the crucial need of accurate low-order descriptions of highly-complex geophysical phenomena and the recent numerical revolution which has supplied the geophysical scientists with an unprecedented volume of image data. The project is placed at the intersection of several fields of expertise (Bayesian inference, matrix factorization, sparse representations, etc.) which will be combined to handle the uncertainties associated to image measurements and to characterize the accurate reduced dynamical systems.

# 9.1.3. ANR BECOSE : Beyond Compressive Sensing: Sparse approximation algorithms for ill-conditioned inverse problems.

Participants: Dominique Heitz, Cédric Herzet.

*duration 48 months.* The BECOSE project aims to extend the scope of sparsity techniques much beyond the academic setting of random and well-conditioned dictionaries. In particular, one goal of the project is to step back from the popular L1-convexification of the sparse representation problem and consider more involved nonconvex formulations, both from a methodological and theoretical point of view. The algorithms will be assessed in the context of tomographic Particle Image Velocimetry (PIV), a rapidly growing imaging technique in fluid mechanics that will have strong impact in several industrial sectors including environment, automotive and aeronautical industries. The consortium gathers the Fluminance and Panama Inria research teams, the Research Center for Automatic Control of Nancy (CRAN), The Research Institute of Communication and Cybernetics of Nantes (IRCCyN), and ONERA, the French Aerospace Lab.

# 9.1.4. ANR-MN: H2MNO4 project

Participants: Yvan Crenner, Benjamin Delfino, Jean-Raynald de Dreuzy, Jocelyne Erhel, Lionel Lenôtre.

Contract with ANR, program Modèles Numériques Duration: four years from November 2012 until April 2017. Title: Original Optimized Object Oriented Numerical Model for Heterogeneous Hydrogeology. Coordination: Jocelyne Erhel and Géraldine Pichot, with Fabienne Cuyollaa. Partners: Geosciences Rennes, University of Poitiers, University of Lyon 1, Andra, Itasca. International collaborations: University of San Diego (USA), UPC, Barcelona (Spain) Web page: http://h2mno4.inria.fr/ Abstract: The project H2MNO4 develops numerical models for reactive transport in heterogeneous media. It defines six mathematical and computational challenges and three applications for environmental problems with societal impact.

# 9.1.5. GDR MANU

Participants: Yvan Crenner, Jocelyne Erhel, Bastien Hamlat.

Title: Mathematics for Nuclear industry Duration: From 2016 to 2019 Coordination: C. Cancès Webpage: http://gdr-manu.math.cnrs.fr/

Abstract: The working group MANU is a follow-up to the group MOMAS. It covers many subjects related to mathematical modeling and numerical simulations for problems arising from nuclear industry and nuclear waste disposal. The team organizes a workshop on reactive transport, Paris, February 2018.

# 9.2. International Initiatives

# 9.2.1. Inria Associate Teams Not Involved in an Inria International Labs

9.2.1.1. LFD-FLU

Title: Large-scale Fluid Dynamics analysis from FLow Uncertainty

International Partner (Institution - Laboratory - Researcher):

Universidad de Buenos Aires (Argentina) - Department of Computer Science and Electrical Engineering - Guillermo Artana

Start year: 2016

See also: http://www.irisa.fr/prive/memin/LFD-FLU/

The first objective of this associate team is primarily concerned with the establishment of efficient fluid flow image data analysis procedures. This concerns for instance data assimilation issues to reconstruct meaningful numerical representation of experimental fluid flows for analysis purpose. The second objective focuses on the incorporation of uncertainties in the flow dynamical evolution models

#### 9.2.1.2. Informal International Partners

**Imperial College**, London (UK), Collaboration with Dan Crisan and Darryl Holm on Stochastic transport for the upper ocean dynamics

**Chico California State University** (USA), We have pursued our collaboration with the group of Shane Mayor on the GPU implementation of wavelet based motion estimator for Lidar data. This code is developped in coproperty between Inria and Chico.

### 9.2.2. Participation in Other International Programs

**Royal Society funding**, collaboration between Dominique Heitz, Etienne Mémin and Sylvain Laizet (Imperial College) on Stochastic large-eddies simulation and data assimilation for the reconstruction of 3D turbulent flows.

**China Scholarship Council funding**, Collaboration between Etienne Memin, Shengze Cai and Chao Xu (Zhejiang University, College of Control Science & Engineering), on turbulent motion estimation and modeling under uncertainty.

# 9.3. International Research Visitors

• 3 weeks visit of Alejandro Gronskis (Researcher Conicet Argentina) to work with Dominique Heitz, Etienne Mémin and Pranav Chandramouli within the associate team LFD

• Sojourn of 12 month of Shengze Cai PhD student in the College of Control Science & Engineering, Zhejiang University to work with Etienne Mémin

# **10.** Dissemination

# **10.1. Promoting Scientific Activities**

# 10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- J. Erhel organizes with J-R. de Dreuzy and T. Le Borgne the international conference CWMR (Saint-Malo, France, June 2018).
- E. Mémin organizes the national colloquium on data assimilation (Rennes, september 2018).

# 10.1.1.2. Member of the Organizing Committees

Jocelyne Erhel

- organizes with J-R. de Dreuzy and T. Le Borgne (OSUR) the international conference CWMR (Saint-Malo, France, June 2018).
- organizes with T. Faney and A. Michel (IFPEN) a workshop on reactive transport (Paris, France, February 2018).

Etienne Mémin

• "Data Science & Environment", workshop + summer school, 3-7 July 2017, Brest, France

# 10.1.2. Scientific Events Selection

# 10.1.2.1. Member of the Conference Program Committees

Jocelyne Erhel

- international advisory committee of the parallel CFD conferences (Glasgow, UK, May 2017).
- program committee of the international conference PARENG 2017 (Pecs, Hungary, May 2017).
- program committee of the workshop Visualization in Environmental Sciences 2017 (co-event of EuroVis)

Cedric Herzet

• Co-organisateur & co-chair of the special session: « Compressed sensing et inversion » of the Gretsi 2017

# 10.1.3. Journal

# 10.1.3.1. Member of the Editorial Boards

Jocelyne Erhel

- member of the editorial board of ETNA.
- member of the editorial board of ESAIM:Proceedings and Surveys.

Etienne Mémin

- Associate editor for the Int. Journal of Computer Vision (IJCV)
- Associate editor for the Image and Vision Computing Journal (IVC)

# 10.1.3.2. Reviewer - Reviewing Activities

Jocelyne Erhel: Reviewer for the journals ADWR, ARIMA, JCAM, MATCOM Dominique Heitz: Reviewer for Exp. in Fluids, SIGGRAPH, ICPMS, Im. Vis. Comp. Cédric Herzet: Reviewer for IEEE Tr. on Signal Processing, IEEE Tr. on Information Theory, Gresti 2017, ICASSP 2017

Etienne Mémin: Reviewer for Tellus-A, Journ. of Fluid Mech., Im. Vis. Comp., Exp. in Fluids, Nonlinear Proc. in Geophysics., Journ. of Comp. Phys.

# 10.1.4. Invited Talks

Dominique Heitz

- 2nd Workshop on Data Assimilation & CFD Processing for PIV and Lagrangian Particle Tracking, Delft, Nederland, 13-14, December, 2017.
- Journée Technique 41- Association Française de Vélomimétrie Laser, Meudon, France, May 11, 2017.

Cédric Herzet

- Séminaire, LIMSI Model Reduction from Partial Observations Jan. 2017
- Séminaire, IMT Atlantique, nov. 2017
- Séminaire, UMR Lab-sticc, Brest, avril 2017

Roger Lewandowski

• Colloquium Univ. Pisa, May 2017.

Etienne Mémin

- AFVL, Meudon, Stratégies d'assimilation de données pour la reconstruction d'écoulements, France, March 2017
- Imperial College, Math. for Planet Earth Centre for doctoral training, Stochastic representation of fluid flow dynamics, November 2017.
- Imperial College, London (UK), Math. for Planet Earth Centre for doctoral training, Stochastic representation of fluid flow dynamics, Nov. 2017.
- Center for Earth System Research and Sustainability, Hamburg (Ger.), Stochastic representation of fluid flow dynamics, Dec. 2017.
- Université de Grenoble, Comodo workshop, Stochastic representation of fluid flow dynamics, Dec. 2017.

# 10.1.5. Leadership within the Scientific Community

• J. Erhel is scientific coordinator of the website Interstices (since June 2012). https://interstices.info.

# 10.1.6. Scientific Expertise

- J. Erhel is a member of the scientific council of IFPEN, since April 2016.
- C. Herzet was reviewer for ANR.

# 10.1.7. Research Administration

Jocelyne Erhel

- correspondent of Maison de la Simulation for Inria Rennes.
- correspondent of AMIES for Inria Rennes, from September 2015.
- member of the Inria national committee for secondment, 2016.
- coordination of the working group for a team creation at Inria Rennes.
- member of the Inria local committee for health and safety (référent chercheur) from January 2016.
- member of the Inria administrative commission (CAP) for researchers, from January 2016.

Dominique Heitz

- Responsible of the Irstea ACTA Team
- Member of Pôle Cristal scientific council
- Member of Irstea OPAALE research unit Executive Committee
- Member of Irstea center of Rennes Executive Committee

Roger Lewandowski

- Elected president of the SMAI-GAMNI thematic group on Advances of Engineering Numerical Methods
- Correspondant SMAI

Etienne Mémin

- Responsible of the "Commission Développement Technologique" Inria Rennes
- Member of the "Commission Personnel" Inria-IRISA Rennes

# **10.2.** Teaching - Supervision - Juries

# 10.2.1. Teaching

Licence: Jocelyne Erhel, Optimisation, 24h, niveau L3, ENSAI Rennes

Licence : Dominique Heitz, Mécanique des fluides, 30h, niveau L2 INSA Rennes

Master: Jocelyne Erhel, arithmétique flottante, 4h, niveau M1, INSA Rennes

Master : Dominique Heitz, Mécanique des fluides, 25h, niveau M1, Dep GMA INSA Rennes

Master : Cédric Herzet, « Smart Sensing », 24h, master international « Smart Sensing and Big Data », ENSAI

Master : Cédric Herzet, Représentations parcimonieuses et compressed sensing, niveau M2, niveau M2, INSA, 10h

Master: Roger Lewandowski, Euler and the Navier-Stokes equations, M2, master « fondamental mathematics ».

Master : Etienne Mémin, Analyse du mouvement, Mastere Informatique, 15h, niveau M2, Université de Rennes 1.

Master : Etienne Mémin, Vision par ordinateur , 15h, niveau M2, ESIR Université de Rennes 1.

# 10.2.2. Supervision

PhD in progress: Pierre-Marie Gibert, University of Lyon, October 2015, co-advisors D. Tromeur-Dervout and Jocelyne Erhel.

PhD in progress: Bastien Hamlat, University of Rennes 1, October 2016, co-advisors Jocelyne Erhel and A. Michel.

PhD in progress : Benoit Pinier, Scale similarity and uncertainty for Ocean-Atmosphere coupled models, started 01/10/2014, supervisors: Roger Lewandowski, Etienne Mémin

PhD in progress : Musaab Khalid Osman Mohammed, Motion analysis techniques for flood images, started february 2015, Lionel Penard (Irstea/Lyon) and Etienne Mémin

PhD in progress : Pranav Chandamouli, Turbulent complex flows reconstruction via data assimilation in large eddy models, started october 2015, Dominique Heitz, Etienne Mémin.

PhD in progress : Romain Schuster, Large-scale fluid motion estimation, started october 2016, Dominique Heitz, Etienne Mémin.

PhD in progress : Long Li, Data assimilation and stochastic transport for the upper ocean dynamics, started november 2017, Etienne Mémin.

PhD in progress : Yacine Ben Ali, Variational assimilation of RANS models for wind engineering, started november 2017, Dominique Heitz, Etienne Mémin.

PhD in progress : Dinh Duong Nguyen, Regular and singular solutions of Navier-Stokes equations with eddy viscosity, started in september 2017, Roger Lewandowski.

# 10.2.3. Juries

Johan Carlier

• Kevin Chatelain, PhD, Polytech'Orléans, (examinateur)

Jocelyne Erhel

- Martin Bachet, PhD, Univ. PSL Mines ParisTech (rapporteur)
- Daniel Jara Heredia, PhD, Univ. Rennes (présidente)
- Antoine Plet, PhD, ENS Lyon (présidente)
- Tangi Migot, PhD, INSA Rennes (examinatrice)

Dominique Heitz

• Robin Yegavian, PhD Ecole Polytechnique (Rapporteur)

Cédric Herzet

• Jean-François Determe PhD université catholique de Louvain, Belgique(Examinateur)

Roger Lewandowski

• Charles Pelletier, PhD Univ. Grenoble Alpes( Rapporteur)

Etienne Mémin

- Olivier Barrois, PhD Univ. Grenoble Alpes( Rapporteur)
- Jean-Matthieu Haussaire, PhD Univ. Paris-Est (Examinateur),

# **10.3.** Popularization

Jocelyne Erhel

- présidente du jury du rallye de mathématiques du CNED, 2017 and 2018.
- First prize of the second Mathematics of Planet Earth international competition. Module "Simulating the melting of ice caps", authors M. Nodet and J. Erhel.

Dominique Heitz

• Interview dans L'Usine Nouvelle, No 3479, pp. 30-31, 2016

# 11. Bibliography

# Major publications by the team in recent years

- [1] K. BURRAGE, J. ERHEL. On the performance of various adaptive preconditioned GMRES, in "Numerical Linear Algebra with Applications", 1998, vol. 5, p. 101-121.
- [2] J. CARRAYROU, J. HOFFMANN, P. KNABNER, S. KRÄUTLE, C. DE DIEULEVEULT, J. ERHEL, J. VAN DER LEE, V. LAGNEAU, K. MAYER, K. MACQUARRIE. Comparison of numerical methods for simulating strongly non-linear and heterogeneous reactive transport problems. The MoMaS benchmark case, in "Computational Geosciences", 2010, vol. 14, n<sup>o</sup> 3, p. 483-502.
- [3] T. CHACÓN-REBOLLO, R. LEWANDOWSKI. *Mathematical and Numerical Foundations of Turbulence Models* and Applications, Modeling and Simulation in Science, Engineering and Technology, Birkhäuser Basel, 2014.
- [4] T. CORPETTI, D. HEITZ, G. ARROYO, E. MÉMIN, A. SANTA-CRUZ. Fluid experimental flow estimation based on an optical-flow scheme, in "Experiments in fluids", 2006, vol. 40, p. 80–97.
- [5] J.-R. DE DREUZY, A. BEAUDOIN, J. ERHEL. Asymptotic dispersion in 2D heterogeneous porous media determined by parallel numerical simulations, in "Water Resource Research", 2007, vol. 43, n<sup>o</sup> W10439, doi:10.1029/2006WR005394.

- [6] A. GRONSKIS, D. HEITZ, E. MÉMIN.Inflow and initial conditions for direct numerical simulation based on adjoint data assimilation, in "Journal of Computational Physics", 2013, vol. 242, p. 480-497 [DOI : 10.1016/J.JCP.2013.01.051], http://www.sciencedirect.com/science/article/pii/ S0021999113001290.
- [7] D. HEITZ, E. MÉMIN, C. SCHNOERR. Variational Fluid Flow Measurements from Image Sequences: Synopsis and Perspectives, in "Experiments in fluids", 2010, vol. 48, n<sup>o</sup> 3, p. 369–393.
- [8] H. HOTEIT, J. ERHEL, R. MOSÉ, B. PHILIPPE, P. ACKERER.*Numerical Reliability for Mixed Methods Applied* to Flow Problems in Porous Media, in "Computational Geosciences", 2002, vol. 6, p. 161-194.
- [9] E. MÉMIN. Fluid flow dynamics under location uncertainty, in "Geophysical & Astrophysical Fluid Dynamics", 2014, vol. 108, n<sup>o</sup> 2, p. 119-146, http://dx.doi.org/10.1080/03091929.2013.836190.
- [10] N. NASSIF, J. ERHEL, B. PHILIPPE. Introduction to computational linear Algebra, CRC Press, 2015.
- [11] N. PAPADAKIS, E. MÉMIN.A variational technique for time consistent tracking of curves and motion, in "Journal of Mathematical Imaging and Vision", 2008, vol. 31, n<sup>o</sup> 1, p. 81–103, http://www.irisa.fr/fluminance/ publi/papers/Papadakis-Memin-JMIV07.pdf.
- [12] Y. SAAD, M. YEUNG, J. ERHEL, F. GUYOMARC'H.A deflated version of the Conjugate Gradient Algorithm, in "SIAM Journal on Scientific Computing", 2000, vol. 21, n<sup>o</sup> 5, p. 1909-1926.

# **Publications of the year**

# **Doctoral Dissertations and Habilitation Theses**

[13] V. RESSEGUIER. *Mixing and fluid dynamics under location uncertainty*, Université Rennes 1, January 2017, https://tel.archives-ouvertes.fr/tel-01507292.

#### **Articles in International Peer-Reviewed Journal**

- [14] S. CAI, E. MÉMIN, P. DÉRIAN, C. XU. Motion Estimation under Location Uncertainty for Turbulent Fluid Flow, in "Experiments in Fluids", 2017, p. 1-17, https://hal.inria.fr/hal-01589642.
- [15] B. CHAPRON, P. DÉRIAN, E. MÉMIN, V. RESSEGUIER. Large scale flows under location uncertainty: a consistent stochastic framework, in "Quarterly Journal of the Royal Meteorological Society", 2017, vol. 00, p. 1 - 15, https://hal.inria.fr/hal-01629898.
- [16] P. DÉRIAN, R. ALMAR. Wavelet-Based Optical Flow Estimation of Instant Surface Currents From Shore-Based and UAV Videos, in "IEEE Transactions on Geoscience and Remote Sensing", 2017, p. 1 - 8 [DOI: 10.1109/TGRS.2017.2714202], https://hal.archives-ouvertes.fr/hal-01573715.
- [17] J. ERHEL, S. SABIT.Analysis of a global reactive transport model and results for the Mo-MaS benchmark, in "Mathematics and Computers in Simulation", 2017, vol. 137, p. 286-298 [DOI: 10.1016/J.MATCOM.2016.11.008], https://hal.inria.fr/hal-01405698.

- [18] P. HÉAS, C. HERZET. Reduced Modeling of Unknown Trajectories, in "Archives of Computational Methods in Engineering", January 2018, vol. 25, n<sup>o</sup> 1, p. 87-101, https://arxiv.org/abs/1702.08846 [DOI: 10.1007/s11831-017-9229-0], https://hal.archives-ouvertes.fr/hal-01490572.
- [19] D. IMBERTI, J. ERHEL. *Vary the s in Your s-step GMRES*, in "Electronic Transactions on Numerical Analysis (ETNA)", 2017, p. 1-27, https://hal.inria.fr/hal-01299652.
- [20] S. KADRI HAROUNA, E. MÉMIN. Stochastic representation of the Reynolds transport theorem: revisiting large-scale modeling, in "Computers and Fluids", August 2017, vol. 156, p. 456-469 [DOI: 10.1016/J.COMPFLUID.2017.08.017], https://hal.inria.fr/hal-01394780.
- [21] J. MARÇAIS, J.-R. DE DREUZY, J. ERHEL.Dynamic coupling of subsurface and seepage flows solved within a regularized partition formulation, in "Advances in Water Resources", 2017, vol. 109, p. 94-105 [DOI: 10.1016/J.ADVWATRES.2017.09.008], https://hal-insu.archives-ouvertes.fr/insu-01586870.
- [22] V. RESSEGUIER, E. MÉMIN, B. CHAPRON. Geophysical flows under location uncertainty, Part I Random transport and general models, in "Geophysical and Astrophysical Fluid Dynamics", April 2017, vol. 111, n<sup>o</sup> 3, p. 149-176, https://arxiv.org/abs/1611.02572 [DOI: 10.1080/03091929.2017.1310210], https://hal. inria.fr/hal-01391420.
- [23] V. RESSEGUIER, E. MÉMIN, B. CHAPRON. Geophysical flows under location uncertainty, Part II Quasigeostrophy and efficient ensemble spreading, in "Geophysical and Astrophysical Fluid Dynamics", April 2017, vol. 111, n<sup>o</sup> 3, p. 177-208 [DOI: 10.1080/03091929.2017.1312101], https://hal.inria.fr/hal-01391476.
- [24] V. RESSEGUIER, E. MÉMIN, B. CHAPRON. Geophysical flows under location uncertainty, Part III SQG and frontal dynamics under strong turbulence conditions, in "Geophysical and Astrophysical Fluid Dynamics", April 2017, vol. 111, n<sup>o</sup> 3, p. 209-227 [DOI: 10.1080/03091929.2017.1312102], https://hal.inria.fr/hal-01391484.
- [25] V. RESSEGUIER, E. MÉMIN, D. HEITZ, B. CHAPRON. Stochastic modelling and diffusion modes for proper orthogonal decomposition models and small-scale flow analysis, in "Journal of Fluid Mechanics", October 2017, vol. 828, 29, https://arxiv.org/abs/1611.06832 [DOI: 10.1017/JFM.2017.467], https://hal.inria.fr/hal-01400119.
- [26] Y. YANG, E. MÉMIN. *High-resolution data assimilation through stochastic subgrid tensor and parameter estimation from 4DEnVar*, in "Tellus A", April 2017, 19, https://hal.inria.fr/hal-01500140.

# **Invited Conferences**

- [27] Y. CRENNER, J. ERHEL. Influence of a fractured medium on pyrite oxidation reaction, in "MAMERN 2017 -International Conference on Approximation Methods and Numerical Modelling in Environment and Natural Resources", Oujda, Morocco, May 2017, https://hal.inria.fr/hal-01646245.
- [28] J. ERHEL.Krylov methods applied to reactive transport models, in "SIAM Conference on Mathematical and Computational Issues in the Geosciences", Erlangen, Germany, September 2017, https://hal.inria.fr/hal-01646268.

# **International Conferences with Proceedings**

- [29] D. ANDA-ONDO, J. CARLIER, C. COLLEWET. Closed-loop control of a spatially developing free shear flow around a steady state, in "20th World Congress of the International Federation of Automatic Control", Toulouse, France, International Federation of Automatic Control, July 2017, https://hal.inria.fr/hal-01514361.
- [30] S. CAI, E. MÉMIN, P. DÉRIAN, C. XU. Location Uncertainty Principle: Toward the Definition of Parameterfree Motion Estimators \*, in "EMMCVPR 2017 - 11th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition", Venice, Italy, Springer, October 2017, p. 1-15, https:// hal.inria.fr/hal-01654184.
- [31] P. CHANDRAMOULI, D. HEITZ, E. MÉMIN, S. LAIZET.A Comparative Study of LES Models Under Location Uncertainty, in "Congrès Français de Mécanique", Lille, France, August 2017, https://hal.inria.fr/ hal-01584736.
- [32] P. CHANDRAMOULI, D. HEITZ, E. MÉMIN, S. LAIZET. Analysis of Models Under Location Uncertainty within the Framework of Coarse Large Eddy Simulation (cLES), in "The 16th European Turbulence Conference", Stockholm, Sweden, European Mechanics Society, August 2017, https://hal.inria.fr/hal-01584735.
- [33] P. HÉAS, C. HERZET. Optimal Low-Rank Dynamic Mode Decomposition, in "Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, New Orleans 2017", New Orleans, United States, March 2017, https://arxiv.org/abs/1701.01064 [DOI : 10.1109/ICASSP.2017.7952999], https://hal.archives-ouvertes.fr/hal-01429975.

# **Conferences without Proceedings**

- [34] R. ALMAR, A. N. LERMA, P. DÉRIAN, B. CASTELLE, T. SCOTT. Transient Surf Zone Circulation Induced by Rhythmic Swash Zone at a Reflective Beach, in "Coastal Dynamics 2017", Helsingør, Denmark, June 2017, https://hal.archives-ouvertes.fr/hal-01573717.
- [35] D. ANDA-ONDO, J. CARLIER, C. COLLEWET. Contrôle en boucle fermée d'une couche de mélange spatiale, in "congrès français de mécanique", Lille, France, August 2017, https://hal.archives-ouvertes.fr/hal-01578477.
- [36] A. GRONSKIS, D. HEITZ, E. MÉMIN. A new hybrid optimization algorithm for variational data assimilation of unsteady wake flows, in "2nd Workshop on Data Assimilation & CFD Processing for Particle Image and Tracking Velocimetry", Delft, Netherlands, December 2017, https://hal.archives-ouvertes.fr/hal-01671776.
- [37] B. HAMLAT, J. ERHEL, A. MICHEL, T. FANEY.*Modélisation des systèmes cinétiques limités*, in "SMAI 2017", La Tremblade, France, June 2017, https://hal.inria.fr/hal-01646320.
- [38] D. IMBERTI, J. ERHEL.Solving large sparse linear systems with a variable s-step GMRES preconditioned by DD, in "DD24 - International Conference on Domain Decomposition Methods", Longyearbyen, Norway, February 2017, https://hal.inria.fr/hal-01528636.
- [39] M. KHALID, L. PÉNARD, E. MÉMIN. Application of optical flow for river velocimetry, in "IGARSS 2017 -37th IEEE Geoscience and Remote Sensing Symposium", Fort Worth, Texas, United States, July 2017, p. 1-4, https://hal.archives-ouvertes.fr/hal-01657202.
- [40] K. ROCHA-BROWNELL, P. DÉRIAN, D. H. RICHTER, P. P. SULLIVAN, S. D. MAYOR. Evaluation of a wavelet-based optical flow algorithm through the use of large eddy simulations, in "28th International Laser Radar Conference", Bucarest, Romania, June 2017, https://hal.archives-ouvertes.fr/hal-01573724.

- [41] R. SCHUSTER, D. HEITZ, E. MÉMIN, A. GUIBERT, P. LOISEL. Stochastic observation model for large scale motion measurement, in "CFM 2017 - 23ème Congrès Français de Mécanique", Lille, France, August 2017, p. 1-12, https://hal.archives-ouvertes.fr/hal-01648441.
- [42] Y. YANG, S. CAI, E. MÉMIN, D. HEITZ. Ensemble-Variational methods in data assimilation, in "2nd Workshop on Data Assimilation & CFD Processing for Particle Image and Tracking Velocimetry", Delft, Netherlands, December 2017, https://hal.archives-ouvertes.fr/hal-01671751.

#### **Research Reports**

[43] S. CAI, E. MÉMIN, Y. YANG, C. XU.Sea Surface Flow Estimation via Ensemble-based Variational Data Assimilation\*, Inria Rennes - Bretagne Atlantique; IRMAR, University of Rennes 1, September 2017, https:// hal.inria.fr/hal-01589637.

# **Other Publications**

- [44] L. C. BERSELLI, R. LEWANDOWSKI. On the Bardina's model in the whole space, 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01590760.
- [45] J. ERHEL, T. MIGOT. Characterizations of Solutions in Geochemistry: Existence, Uniqueness and Precipitation Diagram, September 2017, working paper or preprint, https://hal.inria.fr/hal-01584490.
- [46] P. HÉAS, C. HERZET. Optimal Kernel-Based Dynamic Mode Decomposition, 2017, working paper or preprint [DOI: 10.10919], https://hal.archives-ouvertes.fr/hal-01634542.
- [47] R. LEWANDOWSKI.Navier-Stokes equations in the whole space with an eddy viscosity, May 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01531260.

# **References in notes**

- [48] R. ADRIAN. *Particle imaging techniques for experimental fluid mechanics*, in "Annal Rev. Fluid Mech.", 1991, vol. 23, p. 261-304.
- [49] G. ARTANA, A. CAMMILLERI, J. CARLIER, E. MÉMIN. Strong and weak constraint variational assimilations for reduced order fluid flow modeling, in "J. of Comput. Phys", 2012, vol. 231, n<sup>o</sup> 8, p. 3264–3288.
- [50] T. BEWLEY. Flow control: new challenges for a new Renaissance, in "Progress in Aerospace Sciences", 2001, vol. 37, p. 21–58.
- [51] S. BEYOU, A. CUZOL, S. GORTHI, E. MÉMIN. Weighted Ensemble Transform Kalman Filter for Image Assimilation, in "TellusA", January 2013, vol. 65, n<sup>0</sup> 18803.
- [52] S. CAI, E. MÉMIN, P. DÉRIAN, C. XU. Motion Estimation under Location Uncertainty for Turbulent Fluid Flow, in "Exp. in Fluids", 2018, vol. 59, n<sup>o</sup> 8.
- [53] A. CAMMILLERI, F. GUENIAT, J. CARLIER, L. PASTUR, E. MÉMIN, F. LUSSEYRAN, G. ARTANA.POD-Spectral Decomposition for Fluid Flow Analysis and Model Reduction, in "Theor. and Comp. Fluid Dyn", 2013, Accepted for publication.

- [54] H. CHOI, P. MOIN, J. KIM.Direct numerical simulation of turbulent flow over riblets, in "Journal of Fluid Mechanics", 1993, vol. 255, p. 503–539.
- [55] T. CORPETTI, E. MÉMIN. Stochastic Uncertainty Models for the Luminance Consistency Assumption, in "IEEE Trans. Image Processing", 2012, vol. 21, n<sup>o</sup> 2, p. 481–493.
- [56] G. EVENSEN.Sequential data assimilation with a non linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, in "J. Geophys. Res.", 1994, vol. 99 (C5), n<sup>o</sup> 10, p. 143–162.
- [57] J. FAVIER. Contrôle d'écoulements : approche expérimentale et modélisation de dimension réduite, Institut National Polytechnique de Toulouse, 2007.
- [58] J. FAVIER, A. KOURTA, G. LEPLAT. Control of flow separation on a wing profile using PIV measurements and POD analysis, in "IUTAM Symposium on Flow Control and MEMS", London, UK, September 19-22, 2006.
- [59] N. GORDON, D. SALMOND, A. SMITH. Novel approach to non-linear/non-Gaussian Bayesian state estimation, in "IEEE Processing-F", April 1993, vol. 140, n<sup>O</sup> 2.
- [60] A. GUÉGAN, P. SCHMID, P. HUERRE. *Optimal energy growth and optimal control in swept Hiemenz flow*, in "J. Fluid Mech.", 2006, vol. 566, p. 11–45.
- [61] B. HORN, B. SCHUNCK. Determining Optical Flow, in "Artificial Intelligence", August 1981, vol. 17, n<sup>o</sup> 1-3, p. 185–203.
- [62] F.-X. LE DIMET, O. TALAGRAND. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects, in "Tellus", 1986, n<sup>O</sup> 38A, p. 97–110.
- [63] J. LIONS. Optimal Control of Systems Governed by Partial Differential Equations, Springer-Verlag, 1971.
- [64] L. MATHELIN, O. LE MAÎTRE. Robust control of uncertain cylinder wake flows based on robust reduced order models, in "Computer and Fluids", 2009, vol. 38, p. 1168–1182.
- [65] B. PROTAS, J. WESFREID. Drag force in the open-loop control of the cylinder wake in the laminar regime, in "Physics of Fluids", February 2002, vol. 14, n<sup>o</sup> 2, p. 810–826.
- [66] I. WYGNANSKI. Boundary layer flow control by periodic addition of momentum, in "4th AIAA Shear Flow Control Conference", USA, June 29-July 2, 1997.

# **Project-Team GENSCALE**

# Scalable, Optimized and Parallel Algorithms for Genomics

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: CNRS Université Rennes 1 École normale supérieure de Rennes

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Computational Biology

# **Table of contents**

1.	Personnel	. 341
2.	Overall Objectives	. 342
	2.1. Genomic data processing	342
	2.2. Life science partnerships	342
3.	Research Program	. 343
	3.1. Axis 1: Data Structure	343
	3.2. Axis 2: Algorithms	343
	3.3. Axis 3: Parallelism	344
4.	Application Domains	. 344
	4.1. Introduction	344
	4.2. Health	344
	4.3. Agronomy and Environment	344
5.	Highlights of the Year	. 345
6.	New Software and Platforms	. 345
	6.1. GATB-Core	345
	6.2. DiscoSnpRad	345
	6.3. GWASDM	346
	6.4. bcool	346
_	6.5. CARNAC-LR	346
7.	New Results	. 347
	7.1. Data Structure	347
	7.1.1. Minimal perfect hash function	347
	7.1.2. Quasi-dictionary	347
	7.2. Algorithms & Methods	347
	7.2.1. Short Read Correction	347
	7.2.2. Long transcriptomic read clustering	347
	7.2.3. Statistically Significant Discriminative Patterns Search	348
	7.2.4. Reference free SNP detection in RAD-seq data	348
	7.2.5. Global Optimization for Scaffolding and Completing Genome Assemblies	348
	7.2.6. Identification and characterization of long non-coding RNA	348
	7.2. Derollolier	248 240
	7.5. Parallelisiii 7.4. Disinformatics Analysis	249
	7.4. Diomiormatics Analysis	249
	7.4.1. Study of maine plankton holoolonis	349
	7.4.2. For applied metagenomics	350
	7.4.4 Benchmark of de novo read dataset compression tools	350
	7.4.5 Genomics of the agro-ecosystems pests	350
	7.4.6 Comparison of approaches for finding alternative splicing events in RNA-seq	350
	7.4.7 Microbial communities interaction between plant and their bioagressors	350
	7.5. Challenges	351
8	Bilateral Contracts and Grants with Industry	351
0.	8.1 Bilateral Contracts with Industry	351
	8.2 Bilateral Grants with Industry	351
	8.2.1 Enancio Start-Un	351
	8.2.2. Ransodyn project	351
9.	Partnerships and Cooperations	.352
-•	9.1. Regional Initiatives	352
	9.1.1. Rennes Hospital, Hematology service	352

	9.1.2. Partnership with INRA in Rennes	352
	9.2. National Initiatives	352
	9.2.1. ANR	352
	9.2.1.1. Project HydroGen: Metagenomic applied to ocean life study	352
	9.2.1.2. Project SpeCrep: speciation processes in butterflies	352
	9.2.2. PIA: Programme Investissement d'Avenir	352
	9.2.2.1. RAPSODYN: Optimization of the rapeseed oil content under low nitrogen	352
	9.2.2.2. Institut Français de Bioinformatique: Plant node	353
	9.2.3. Programs from research institutions	353
	9.2.3.1. Inria ADT DiagCancer	353
	9.2.3.2. CNRS Mastodons program: C3G	353
	9.2.3.3. Inria Project Lab: Neuromarkers	353
	9.3. International Initiatives	353
	9.3.1. Inria Associate Team: HipcoGen	353
	9.3.2. Informal International Partners	354
	9.4. International Research Visitors	354
	9.4.1. Visits of International Scientists	354
	9.4.2. Visits to International Teams	354
10.	Dissemination	354
	10.1. Promoting Scientific Activities	354
	10.1.1. Scientific Events Organisation	354
	10.1.2. Scientific Events Selection	354
	10.1.2.1. Member of the Conference Program Committees	354
	10.1.2.2. Reviewer	354
	10.1.3. Journal	354
	10.1.4. Invited Talks	354
	10.1.5. Leadership within the Scientific Community	355
	10.1.6. Scientific Expertise	355
	10.1.7. Research Administration	355
	10.2. Teaching - Supervision - Juries	355
	10.2.1. Teaching	355
	10.2.2. Supervision	356
	10.2.3. Juries	356
	10.3. Popularization	356
11.	Bibliography	. 356

# **Project-Team GENSCALE**

*Creation of the Team: 2012 January 01, updated into Project-Team: 2013 January 01* **Keywords:** 

# **Computer Science and Digital Science:**

A3.1.2. - Data management, quering and storage

A3.1.8. - Big data (production, storage, transfer)

A3.3.2. - Data mining

A3.3.3. - Big data analysis

A7.1. - Algorithms

A8.2. - Optimization

# **Other Research Topics and Application Domains:**

B1.1.6. - Genomics

B1.1.9. - Bioinformatics

B2.2.3. - Cancer

# 1. Personnel

# **Research Scientists**

Dominique Lavenier [Team leader, CNRS, Senior Researcher, HDR] Claire Lemaitre [Inria, Researcher] Pierre Peterlongo [Inria, Researcher, HDR]

#### **Faculty Member**

Rumen Andonov [Univ de Rennes I, Professor, HDR]

# **External Collaborators**

Susete Alves Carvalho [INRA, from Mar 2017] Fabrice Legeai [INRA] Guillaume Rizk [AlgoRizk Company]

### **Technical Staff**

Jennifer Del Giudice [Inria, until Mar 2017] Patrick Durand [Inria, until Aug 2017] Jeremy Gauthier [Inria, from Feb 2017] Sebastien Letort [CNRS] Stephane Picq [Inria, from Feb 2017 until Jun 2017] Charles Deltel [Inria, 50%]

# **PhD Students**

Gaetan Benoit [Inria] Wesley Delage [Inria, from Oct 2017] Sebastien Francois [Univ de Rennes I] Cervin Guyomar [Univ de Rennes I] Lolita Lecompte [Inria, from Sep 2017] Antoine Limasset [Univ de Rennes I, until Aug 2017] Camille Marchet [Univ de Rennes I] Hoang Son Pham [Vietnam gov.]

#### Visiting Scientists

Hristo Djidjev [Los Alamos National Laboratory, from Jun 2017 until Jul 2017] Metodi Traykov [Agence Erasmus+ France, from Mar 2017 until Apr 2017]

#### Administrative Assistant

Marie Le Roic [Univ de Rennes I]

# 2. Overall Objectives

# 2.1. Genomic data processing

The main goal of the GenScale project is to develop scalable methods, tools, and software for processing genomic data. Our research is motivated by the fast development of next-generation sequencing (NGS) technologies that provide very challenging problems both in terms of bioinformatics and computer sciences. As a matter of fact, the last sequencing machines generate Tera bytes of DNA sequences from which time-consuming processes must be applied to extract useful and pertinent information.

Today, a large number of biological questions can be investigated using genomic data. DNA is extracted from one or several living organisms, sequenced with high throughtput sequencing machines, then analyzed with bioinformatics pipelines. Such pipelines are generally made of several steps. The first step performs basic operations such as quality control and data cleaning. The next steps operate more complicated tasks such as genome assembly, variant discovery (SNP, structural variations), automatic annotation, sequence comparison, etc. The final steps, based on more comprehensive data extracted from the previous ones, go toward interpretation, generally by adding different semantic information, or by performing high-level processing on these pre-processed data.

GenScale expertise relies mostly on the first and second steps. The challenge is to develop scalable algorithms able to devour the daily DNA flow that tends to congest the bioinformatics computing centers. To achieve this goal, our strategy is to work both on space and time scalability aspects. Space scalability is correlated to the design of optimized and low memory footprint data structures able to capture all useful information contained in sequencing datasets. The idea is that hundreds of Giga bytes of raw data absolutely need to be represented in a very concise way in order to completely fit into a computer memory. Time scalability means that the execution of the algorithms must be as short as possible or, at least, must last a reasonable amount of time. In that case, conventional algorithms that were working on rather small datasets must be revisited to scale on today NGS data. Parallelism is a complementary technique for increasing scalability.

GenScale research is then organized along three main axes:

- Axis 1: Data structures
- Axis 2: Algorithms
- Axis 3: Parallelism

The first axis aims at developing advanced data structures dedicated to sequencing data. Based on these objects, the second axis provides low memory footprint algorithms for a large panel of usual NGS tools. Fast execution time is improved by the third axis. The combination of these three components allows efficient and scalable algorithms to be designed.

# 2.2. Life science partnerships

A second important objective of GenScale is to create and maintain permanent partnerships with other life science research groups. As a matter of fact, the collaboration with genomic research teams is of crucial importance for validating our tools, and for capturing new trends in the bioinformatics domain. Our approach is to actively participate in solving biological problems (with our partners) and to get involved in a few challenging genomic projects.

Partnerships are mainly supported by collaborative projects (such as ANR projects) in which we act as bioinformatics partners either for bringing our expertise in that domain or for developing ad hoc tools.

# 3. Research Program

# 3.1. Axis 1: Data Structure

The aim of this axis is to develop efficient data structures for representing the mass of genomic data generated by the sequencing machines. This research is motivated by the fact that the treatments of large genomes, such as mammalian or plant genomes, require high computing resources, and more specifically very important memory configuration. For example, the ABYSS software used 4.3TB of memory to assemble the white spruce genome [45]. The main reason for such memory consumption is that the data structures used in ABYSS are far from optimal (and this is also the case for many assembly software).

Our research focuses on the de-Bruijn graph structure. This well-known data structure, directly built from raw sequencing data, have many properties matching perfectly well with NGS processing requirements (see next section). Here, the question we are interested in is how to provide a low memory footprint implementation of the de-Bruijn graph to process very large NGS datasets, including metagenomic ones.

Another research direction of this axis is the indexing of large sets of objects. A typical, but non exclusive, need is to annotate nodes of the de-Bruijn graph, that is potentially billions of items. Again, very low memory footprint indexing structures are mandatory to manage a very large quantity of objects.

# 3.2. Axis 2: Algorithms

The main goal of the GenScale team is to develop optimized tools dedicated to NGS processing. Optimization can be seen both in terms of space (low memory footprint) and in terms of time (fast execution time). The first point is mainly related to advanced data structures as presented in the previous section (axis 1). The second point relies on new algorithms and, when possible implementation on parallel structures (axis 3).

We do not have the ambition to cover the vast panel of software related to NGS needs. We particularly focused on the following areas:

- NGS data Compression De-Bruijn graphs are de facto a compressed representation of the NGS information from which very efficient and specific compressors can be designed. Furthermore, compressing the data using smart structures may speed up some downstream graph-based analyses since a graph structure is already built.
- Genome assembly This task remains very complicated, especially for large and complex genomes, such as plant genomes with polyploid and highly repeated structures. We worked both on the generation of contigs and on the scaffolding step.
- **Detection of variants** This is often the first information we want to extract from billions of reads. Variant structures range from SNPs or short indels to large insertions/deletions and long inversions over the chromosomes. We developed original methods to find variants without any reference genome.
- **Metagenomics** We focussed our research on comparative metagenomics by providing methods able to compare hundreds of metagenomic samples together. This is achieved by combining very low memory data structures and efficient implementation and parallelization on large clusters.
- Genome Wide Association Study (GWAS) We tackle this problem with algorithms commonly used in data mining. From two cohorts of individuals (case and control) we can exhibit statistically significant *patterns* spaning over full genomes.

# 3.3. Axis 3: Parallelism

This third axis is another lever to increase performances and scalability of NGS treatments. There are many levels of parallelism that can be used and/or combined to reduce the execution time of very time-consuming bioinformatics processes. A first level is the parallel nature of today processors that now house several cores. A second level is the grid structure that is present in all bioinformatics centers or in the cloud. This two levels are generally combined: a node of a grid is often a multicore system. Another possibility is to add hardware accelerators to a processor. A GPU board is a good example.

GenScale does not do explicit research on parallelism. It exploits the capacity of computing resources to support parallelism. The problem is addressed in two different directions. The first is an engineering approach that uses existing parallel tools to implement algorithms such as multithreading or MapReduce techniques. The second is a parallel algorithmic approach: during the development step, the algorithms are constrained by parallel criteria. This is particularly true for parallel algorithms targeting hardware accelerators.

# 4. Application Domains

# 4.1. Introduction

Today, sequencing data are intensively used in many life science projects. The methodologies developed by the GenScale group are generic approaches that can be applied to a large panel of domains such as health, agronomy or environment areas. The next sections briefly describe examples of our activity in these different domains.

# 4.2. Health

**Genetic and cancer disease diagnostic:** Genetic diseases are caused by some particular mutations in the genomes that alter important cell processes. Similarly, cancer comes from changes in the DNA molecules that alter cell behavior, causing uncontrollable growth and malignancy. Pointing out genes with mutations helps in identifying the disease and in prescribing the right drug. Thus, DNA from individual patients is sequenced and the aim is to detect potential mutations that may be linked to the patient disease. Today the bioinformatics analysis is mainly based on the detection of SNPs (Single Nucleotide Polymorphism) from a set of predefined target genes. Tomorrow, due to the decreasing cost of the sequencing process, bioinformatics analysis will scan the complete genome and report all kinds of mutations, including complex mutations such as large insertions or deletions, that could be associated with cancers.

**Neurodegenerative disorders:** The biological processes that lead from abnormal protein accumulation to neuronal loss and cognitive dysfunction is not fully understood. In this context, neuroimaging biomarkers and statistical methods to study large datasets play a pivotal role to better understand the pathophysiology of neurodegenerative disorders. The discovery of new anatomical biomarkers could thus have a major impact on clinical trials by allowing inclusion of patients at a very early stage, at which treatments are the most likely to be effective. Correlations with genetic variables can determine subgroups of patients with common anatomical and genetic characteristics.

# 4.3. Agronomy and Environment

**Improving plant breeding:** such projects aim at 1) identifying favorable alleles at loci contributing to phenotypic variation, 2) characterizing N-traits at the functional level and 3) providing robust multi-locus SNP-based predictors of the breeding value of agronomical traits under polygenic control. Underlying bioinformatics processing is the detection of informative zones (QTL) on the plant genomes.

**Insect genomics:** Insects represent major crop pests, justifying the need for control strategies to limit population outbreaks and the dissemination of plant viruses they frequently transmit. Several issues are investigated through the analysis and comparison of their genomes: understanding their phenotypic plasticity such as their reproduction mode changes, identifying the genomic sources of adaptation to their host plant and of ecological speciation, and understanding the relationships with their bacterial symbiotic communities.

**Ocean biodiversity:** The metagenomic analysis of seawater samples provides an original way to study the ecosystems of the oceans. Through the biodiversity analysis of different ocean spots, many biological questions can be addressed, such as the plankton biodiversity and their role, for example, in the CO2 sequestration.

# 5. Highlights of the Year

# **5.1. CAMI**

GenScale participated to the international CAMI challenge. CAMI stands for Critical Assessment of Metagenome Interpretation. It is a community-led initiative designed to tackle the problem of recovering the complex information encoded in metagenomes by aiming for an independent, comprehensive and bias-free evaluation of methods. We contributed in the "Assembly" section with the Minia pipeline. Results of this competition, presented in the "Nature Methods" journal [20], highlight the good behaviour of our tool compared to other competitors.

# 6. New Software and Platforms

# 6.1. GATB-Core

# Genome Assembly and Analysis Tool Box

KEYWORDS: Bioinformatics - NGS - Genomics - Genome assembling

FUNCTIONAL DESCRIPTION: The GATB-Core library aims to lighten the design of NGS algorithms. It offers a panel of high-level optimized building blocks to speed-up the development of NGS tools related to genome assembly and/or genome analysis. The underlying data structure is the de Bruijn graph, and the general parallelism model is multithreading. The GATB library targets standard computing resources such as current multicore processor (laptop computer, small server) with a few GB of memory. From high-level API, NGS programming designers can rapidly elaborate their own software based on domain state-of-the-art algorithms and data structures. The GATB-Core library is written in C++.

RELEASE FUNCTIONAL DESCRIPTION: speed up from x2 to x4 for kmer counting and graph construction phases (optimizations based on minimizers and improved Bloom filters). GATB's k-mer counter has been improved using techniques from KMC2, to achieve competitive running times compared to KMC2. ability to store arbitrary information associated to each kmer of the graph, enabled by a minimal perfect hash function (costs only 2.61 bits/kmer of memory) improved API with new possibilities (banks and kmers management) many new snippets showing how to use the library.

- Participants: Charles Deltel, Claire Lemaitre, Dominique Lavenier, Guillaume Rizk, Patrick Durand and Pierre Peterlongo
- Contact: Dominique Lavenier
- URL: http://gatb.inria.fr/

# 6.2. DiscoSnpRad

DISCOvering Single Nucleotide Polymorphism, Indels in RAD seq data KEYWORD: RAD-seq FUNCTIONAL DESCRIPTION: Software discoSnpRad is designed for discovering Single Nucleotide Polymorphism (SNP) and insertions/deletions (indels) from raw set(s) of RAD-seq data. Note that number of input read sets is not constrained, it can be one, two, or more. Note also that no other data as reference genome or annotations are needed. The software is composed of several modules. First module, kissnp2, detects SNPs from read sets. A second module, kissreads2, enhances the kissnp2 results by computing per read set and for each variant found i/ its mean read coverage and ii/ the (phred) quality of reads generating the polymorphism. Then, variants are grouped by RAD locus, and a VCF file is finally generated. We also provide several scripts to further filter and select informative variants for downstream population genetics studies.

This tool relies on the GATB-Core library.

- Contact: Pierre Peterlongo
- URL: https://github.com/GATB/DiscoSnp

# 6.3. GWASDM

Genome Wide Association Study using Data Mining strategy

KEYWORDS: GWAS - Data mining

FUNCTIONAL DESCRIPTION: From two cohorts of genotyped individuals (case and control), the GWASDM software performs a Genome Wide Association Study based on data mining techniques and generates several patterns of SNPs that correlate with a given phenotype. The algorithm implemented in GWASDM directly uses relative risk measures such as risk ratio, odds ratio and absolute risk reduction combined with confidence intervals as anti-monotonic properties to efficiently prune the search space. The algorithm discovers a complete set of discriminating patterns with regard to given thresholds or applies heuristic strategies to extract the largest statistically significant discriminating patterns in a given dataset.

• Contact: Dominique Lavenier

# 6.4. bcool

de Bruijn graph cOrrectiOn from graph aLignment

KEYWORDS: De Bruijn graphs - Reads correction - Short reads - Read mapping

FUNCTIONAL DESCRIPTION: BCool includes two steps. As a first step, Bcool constructs a corrected compacted de Bruijn graph from the reads. This graph is then used as a reference and the reads are corrected according to their mapping on the graph. This approach yields a better correction than kmer-spectrum techniques, while being scalable, making it possible to apply it to human-size genomic datasets and beyond. The implementation is open source and available at github.com/Malfoy/BCOOL

- Partner: Université libre de Bruxelles
- Contact: Pierre Peterlongo
- URL: http://github.com/Malfoy/BCOOL

# 6.5. CARNAC-LR

#### Clustering coefficient-based Acquisition of RNA Communities in Long Reads

**KEYWORDS:** Transcriptomics - Clustering - Bioinformatics

FUNCTIONAL DESCRIPTION: Carnac-LR is a clustering method for third generation sequencing data. Used on RNA sequences it retrieves all sequences that describes a gene and put them in a cluster. CARNAC-LR is an efficient implementation of a novel clustering algorithm for detecting communities in a graph of reads from Third Generation Sequencing. It is a part of a pipeline that allows to retrieve expressed variants from each gene de novo (without reference genome/transcriptome), for transcriptomic sequencing data.

• Contact: Camille Marchet

# 7. New Results

# 7.1. Data Structure

# 7.1.1. Minimal perfect hash function

Participants: Antoine Limasset, Guillaume Rizk, Pierre Peterlongo.

Minimal perfect hash functions are fundamental objects used in many applications. Existing algorithms and implementations that build such functions have in practice some upper bounds on the number of input elements they can handle, due to high construction time and/or memory usage. We propose a simple algorithm having very competitive construction times, memory usage and query times compared to state of the art techniques [27]. We provide a parallel implementation called BBHash. It is capable of creating a minimal perfect hash function of  $10^{10}$  elements in less than 1 hour and 4 GB of memory. To the best of our knowledge, this library is also the first that has been successfully tested on  $10^{12}$  input elements. Source code: https://github.com/rizkg/BBHash

# 7.1.2. Quasi-dictionary

Participants: Camille Marchet, Antoine Limasset, Pierre Peterlongo.

Indexing massive data sets is extremely expensive for large scale problems. In many fields, huge amounts of data are currently generated, however extracting meaningful information from voluminous data sets, such as computing similarity between elements, is far from being trivial. It remains nonetheless a fundamental need. In this context, we proposed a probabilistic data structure based on a minimal perfect hash function for indexing large sets of keys. This structure out-competes the hash table for construction, query times and for memory usage, in the case of the indexation of a static set. To illustrate the impact of algorithms performances, we provided two applications based on similarity computation between collections of sequences, and for which this calculation is an expensive but required operation. In particular, we showed a practical case in which other bioinformatics tools failed to scale up the tested data set or provide lower recall quality results [43].

# 7.2. Algorithms & Methods

# 7.2.1. Short Read Correction

Participants: Antoine Limasset, Pierre Peterlongo.

We proposed a new method to correct short reads using de Bruijn graphs, and we implemented it as a tool called Bcool. As a first step, Bcool constructs a corrected compacted de Bruijn graph from the reads. This graph is then used as a reference and the reads are corrected according to their mapping on the graph. We showed that this approach yields a better correction than kmer-spectrum techniques, while being scalable, making it possible to apply it to human-size genomic datasets and beyond [41].

# 7.2.2. Long transcriptomic read clustering

Participants: Camille Marchet, Pierre Peterlongo.

This contribution tackles the problem of clustering RNA reads in clusters representing all variants of each gene, in a *de novo* way i.e. without any reference sequences. Such problem is not new as is, but the latest, Third Generation Sequencing (TGS) data redefine it. Reads can now span full-length transcripts but at the price of very high error rates, mostly insertions and deletions. This makes difficult or impossible to use tools designed for previous sequencing data. Still, the property to obtain whole RNA molecules through reads is very promising to better describe a transcriptome. In this work, we targeted the need to extract relevant information from a TGS transcriptome, even when no reference is available. In collaboration with Jacques Nicolas from the Inria/IRISA Dyliss team, we propose a novel algorithm in the community detection framework, based on the clustering coefficient. In addition we propose an implementation of this algorithm in the tool CARNAC-LR and a pipeline for the processing of transcriptome data. We validated our tool on real data from mouse and showed that it could be accurate and precise even for lowly expressed genes. We showed that our approach can be complementary to a mapping in the case a reference exists, and that a straightforward use of CARNAC-LR enables to quickly assess the genes'e expression levels [42].

# 7.2.3. Statistically Significant Discriminative Patterns Search

Participants: Hoang Son Pham, Dominique Lavenier.

Identifying multiple SNPs combinations associated with diseases such as cancers or diabetes is a central goal of human genetics. Recently, discriminative pattern mining algorithms have been investigated to tackle genome-wide association studies (GWAS). We designed an algorithm, called SSDPS, to discover groups of items which have significant difference of frequency in case-control datasets. The algorithm directly uses relative risk measures such as risk ratio, odds ratio and absolute risk reduction combined with confidence intervals as anti-monotonic properties to efficiently prune the search space. The algorithm discovers a complete set of discriminative patterns with regard to given thresholds or applies heuristic strategies to extract the largest statistically significant discriminative patterns in a given dataset. Experimental results on both synthetic datasets and three real variant datasets (Age-Related Macular Degeneration, Breast Cancer and Type 2 Diabetes) demonstrate that the SSDPS algorithm effectively detects multiple SNPs combinations in an acceptable execution time.

# 7.2.4. Reference free SNP detection in RAD-seq data

Participants: Jeremy Gauthier, Claire Lemaitre, Pierre Peterlongo.

We developed an original method for reference-free variant calling from Restriction site associated DNA Sequencing (RAD-Seq) data. RAD-seq is a technique widely employed in the evolutionary biology field. Based on the variant caller DiscoSnp, DiscoSnp-RAD explores the De Bruijn Graph built from all the read datasets to detect SNP and Indels. Tested on simulated and real datasets, DiscoSnp-RAD identifies thousands of variants suitable for different population genomics analyses. Furthermore, DiscoSnp-RAD stands out from other tools due to his completely different principle, making it significantly faster, in particular on large datasets [39].

# 7.2.5. Global Optimization for Scaffolding and Completing Genome Assemblies

Participants: Sebastien Francois, Rumen Andonov, Dominique Lavenier.

We developed a method for solving genome scaffolding as a problem of finding the longest simple path in a graph defined by the contigs that satisfies a maximal number of additional constraints encoding the insert-size information [26]. Then we solved the resulting mixed integer linear program to optimality using the Gurobi solver. We tested our algorithm on a benchmark of chloroplast genomes and showed that it outperforms other widely-used assembly solvers by the accuracy of the results.

# 7.2.6. Identification and characterization of long non-coding RNA

# Participant: Fabrice Legeai.

We participated in the development and validation of the tool FeelNC (collaboration with IGDR group). This is a tool allowing the identification of long non coding RNA (lncRNA) from RNASeq reads with or without a reference genome. Contrary to other tools, it does not depend on the comparison with protein databanks, which usually require lots of computations, but used a machine learning approach based on a Random Forest model trained with general features such as multi k-mer frequencies and relaxed open reading frames. We delivered a module that allows to characterize the relationships of each long non coding RNA with the other genes in its genomics close environment, giving insights about the putative impact of the lncRNAs to the regulation of these genes [23], [24].

# 7.2.7. Characterizing repeat-associated subgraphs in de Bruijn graphs

Participant: Camille Marchet.

The main problem in genome assembly, namely repeats, is also present in transcriptomic data. They are dealt with using various heuristics in the de Bruijn Graph framework (dBG). In this work, we introduce a formal model for representing high copy-number and low-divergence repeats in RNA-seq data in dBG and infer the definition of repeat-associated subgraphs. We show that the problem of identifying such subgraphs in a dBG is NP-complete. Then we place ourselves in the case of local assembly of alternative splicing and show that

such subgraphs can be avoided implicitly. Thus, more alternative splicing events can be enumerated than with previous approaches. Finally we show that this exploration of DBG explorations can improve de novo transcriptome evaluation methods [16].

# 7.3. Parallelism

# 7.3.1. Variant detection using processing-in-memory technology

Participants: Charles Deltel, Dominique Lavenier.

The concept of Processing-In-Memory aims to dispatch the computer power near the data. Together with the UPMEM company (http://www.upmem.com/), which is currently developing a DRAM memory enhanced with computing units, we investigate the parallelization of the detection of mutations on the human genome. Traditionnaly, this process is split into 2 steps: a mapping step and a variant calling step. Here, thanks to the high processing power of this new type of memory, the mapping step can nearly be done at the disk transfer rate, allowing the variant calling step to be done simultaneously on the host processor. The implementation is currently on going. First performance evaluations indicate speed-up of one or two order of magnitude compared to purely software implementation.

# 7.4. Bioinformatics Analysis

# 7.4.1. Study of marine plankton holobionts

# Participants: Camille Marchet, Pierre Peterlongo.

We derived from the quasi-dictionary (described in previous section) a tool called Short Read Connector (SRC), able to find pairs of similar reads intra or inter read sets. We used SRC in meta-transcriptomics context to identify the actors of a symbiosis and help the assembly [44], [31]. The framework is the study of marina holobionts (host and its community of symbionts) for which few is known about the actors. In order to retrieve the functions that characterize such holobionts, RNA-seq reads from the sequencing of the whole holobiont are assembled *de novo*. Such assembly is prone to produce chimeras. Thus SRC is used to index sequences (reads, EST, assembled genes...) known to be close to the host and symbionts of the holobiont. Then, thanks to SRC's ability to find similarity between sequences even at a large scale, by querying reads of the holobiont we identify those similar to the host or symbionts. We report four categories: host, symbiont, shared and unassigned that can be assembled in a parallel way. As a first step we validate the SRC+assembly approach by comparing our result to literature with two known holobionts with eukaryote hosts (*Orbicella faveolata, Xestospongia muta*). We show that our approach can compare to previous results. In a second step we lean on a protist (Collodaria) holobiont for which the actors are poorly known. No assembled sequences exist in the literature so we compare the pipeline SRC+assembly to a sole assembly pipeline. Our main achievement is to highly reduce (up to  $\sim$ 40%) the number of chimeras in the assembly compared to the sole assembly pipeline.

# 7.4.2. Pea aphid metagenomics

Participants: Cervin Guyomar, Fabrice Legeai, Claire Lemaitre.

We worked on a framework adapted to the study of genomic diversity and evolutionary dynamics of the pea aphid symbiotic community from an extensive set of metagenomics datasets. The framework is based on mapping to reference genomes and whole genome SNP-calling. We explored the genotypic diversity associated to the different symbionts of the pea aphid at several scales : across host biotypes, amongst individuals of the same biotype, or within individual aphids. Thorough phylogenomic analyses highlighted that the evolutionary dynamics of symbiotic associations strongly varied depending on the symbiont, reflecting different histories and possible constraints [40], [30].

# 7.4.3. Assembly and comparison of two genomes of highly polyphagous lepidopteran pests

Participants: Fabrice Legeai, Claire Lemaitre.

In this study, two genomes of an agronomical important lepidopteran pest, the noctuid moth *Spodoptera frugiperda*, were sequenced and compared, giving significant insights to the mechanisms involved in hostplant adaptation and speciation of this organism. In particular, we described the large expansion of gustatory receptors and detoxification genes among this polyphagous pest compared to other specialist Lepidoptera, and emphasizes the role of these 2 gene families in the evolution of one of the world's worst agricultural pests. We also provided the genome assemblies, gene annotations and whole genome alignments of both strains, and the comparison of both to a reference moth genome (*Bombyx mori*). For these purposes, several original methods were developed i) to correct genome assembly errors due to the high level of heterozygosity and ii) to extract structural variant calls from whole genome alignments [15].

### 7.4.4. Benchmark of de novo read dataset compression tools

Participants: Gaetan Benoit, Dominique Lavenier, Claire Lemaitre.

In this book chapter, we review the different approaches and their tools developed so far to compress sequencing data files. We detail the algorithms for each of the three main types of data contained in such files for each read : the header, the DNA and the quality sequences. We also provide a thorough benchmark of the numerous available tools on various sequencing datasets, evaluating the compression ratio as well as the running time and memory usage performances [33].

# 7.4.5. Genomics of the agro-ecosystems pests

Participants: Fabrice Legeai, Claire Lemaitre.

Within a large international network of biologists, GenScale has contributed to various projects for identifying important components such as protein coding or non coding genes involved in the adaptation of major agricultural pests to their environment. We provided or participated to the assembly and the annotation of 4 new aphids [17], [22], and 5 parasitic wasps. Following specific agreement or policy, these new genomes and annotations are available for a restricted consortium or a large community through the BioInformatics platform for Agro-ecosystems Arthropods (http://bipaa.genouest.org/is). Moreover our engagement in the agronomical pest genomics led to our contribution to other projects such as epigenetics and chromatin structure analysis [18], or the analysis of population genetics data for identifying hotspots of selection in the nematode *Globodera pallida* genome [14].

# 7.4.6. Comparison of approaches for finding alternative splicing events in RNA-seq Participant: Camille Marchet.

In this work we compared an assembly-first and a mapping-first approach to analyze RNA-seq data and find alternative splicing (AS) events. Assembly-first approach enables to identify novel AS events and to detect events in paralog genes that are hard to find using mapping because of the multi-mapping results. On the other hand, the mapping-first approach is more sensitive and detects AS events in lowly expressed genes, and is also able to find AS events with exons containing transposable elements. In addition we support these results with experimental validation. We showed that in order to extensively study the alternative splicing via RNA-seq data and retrieve the most candidates, both approaches should be led. We provide a pipeline constituted of parallel local *de novo* assembly executed by KisSplice and mapping using a novel mapping workflow called FaRLine [37].

### 7.4.7. Microbial communities interaction between plant and their bioagressors

**Participants:** Susete Alves Carvalho, Fabrice Legeai, Claire Lemaitre, Pierre Peterlongo, Dominique Lavenier.

GenScale actively collaborates with the INRA group 'plant-microbial communities interactions' (IGEPP, Rennes) that analyze the interaction between plant, their associated microbial communities and different bioagressors. The ambition of the project is to understand the link between the taxonomic biodiversity of the microbiota and their functional diversity in relation with plant physiology and plant-bioagressors interactions. For this last point, an integrated metatranscritomic approach is developped. Beside wet lab and sequence productions, bioinformatics tools are needed and meta-transcriptomic pipelines analysis arecurrently developped based on the GenScale expertise.

# 7.5. Challenges

# 7.5.1. Participation to CAMI: de-novo metagenomics assembly competition

Participants: Charles Deltel, Dominique Lavenier, Claire Lemaitre, Pierre Peterlongo.

In metagenome analysis, computational methods for assembly, taxonomic profiling and binning are key components facilitating downstream biological data interpretation. However, a lack of consensus about benchmarking datasets and evaluation metrics complicates proper performance assessment. In this context, we participated to CAMI (Critical Assessment of Metagenome Interpretation), specifically on the assembly section with the Minia pipeline. The CAMI challenge aimed to benchmark programs on datasets of unprecedented complexity and realism. Benchmark metagenomes were generated from 700 newly sequenced microorganisms and 600 novel viruses and plasmids, including genomes with varying degrees of relatedness to each other and to publicly available ones and representing common experimental setups. Across all datasets, our assembly programs performed well for species represented by individual genomes, while performance was substantially affected by the presence of related strains [20].

# 8. Bilateral Contracts and Grants with Industry

# 8.1. Bilateral Contracts with Industry

#### 8.1.1. Processing in memory

Participants: Charles Deltel, Dominique Lavenier.

The UPMEM company is currently developing new memory devices with embedded computing power (http:// www.upmem.com/). GenScale investigates how bioinformatics algorithms can benefit from these new types of memory (see section New Results).

# 8.2. Bilateral Grants with Industry

# 8.2.1. Enancio Start-Up

Participants: Jennifer Del Giudice, Stephane Picq, Guillaume Rizk.

After 2 years of development the EnginesOn project has led to the creation of Enancio in August 2017 (http://www.enancio.fr). Enancio main focus is to give the biologist all the resources needed to decipher the information held on a biological molecule such as DNA, without worrying about the informatics behind it. The start-up provides a software platform available through the net with analysis workflows that have been conceived and validated by the field experts, solutions to handle massive data, and health data certified computational infrastructure. Simplification, optimization and faster execution of analyses workflows are the main focuses of the company. Enancio workflows uses the GATB-core library developed by GenScale.

### 8.2.2. Rapsodyn project

Participants: Dominique Lavenier, Claire Lemaitre, Sebastien Letort, Pierre Peterlongo.

RAPSODYN is a long term project funded by the IA ANR French program (Investissement d'Avenir) and several field seed companies, such as Biogemma, Limagrain and Euralis (http://www.rapsodyn.fr/). The objective is the optimization of the rapeseed oil content and yield under low nitrogen input. GenScale is involved in the bioinformatics work package, in collaboration with Biogemma's bioinformatics team, to elaborate advanced tools dedicated to polymorphism detection and analysis.

# 9. Partnerships and Cooperations

# 9.1. Regional Initiatives

# 9.1.1. Rennes Hospital, Hematology service

Participants: Dominique Lavenier, Patrick Durand.

The collaboration with the Hematology service of the Rennes hospital aims to set up advanced bioinformatics pipelines for cancer diagnosis. More precisely, we evaluated a new method of predictions of small cancer-related mutations (such as SNPs and small insertions/deletions) from raw DNA sequencing data.

# 9.1.2. Partnership with INRA in Rennes

**Participants:** Susete Alves Carvalho, Cervin Guyomar, Dominique Lavenier, Fabrice Legeai, Claire Lemaitre, Sebastien Letort, Pierre Peterlongo.

The GenScale team has a strong and long term collaboration with biologists of INRA in Rennes: IGEPP and PEGASE units. This partnership concerns both service and research activities and is acted by the hosting of two INRA engineer (F. Legeai, S. Alves Carvalho) and one PhD student (C. Guyomar).

# 9.2. National Initiatives

# 9.2.1. ANR

9.2.1.1. Project HydroGen: Metagenomic applied to ocean life study

Participants: Dominique Lavenier, Pierre Peterlongo, Claire Lemaitre, Guillaume Rizk, Gaetan Benoit.

Coordinator: P. Peterlongo (Inria/Irisa, GenScale, Rennes) Duration: 42 months (Nov. 2014 – Apr. 2018) Partners: CEA (GenosScope, Evry), INRA (AgroParisTech, Paris – MIG, Jouy-en-Jossas).

The HydroGen project aims to design new statistical and computational tools to measure and analyze biodiversity through comparative metagenomic approaches. The support application is the study of ocean biodiversity based on the analysis of seawater samples available from the Tara Oceans expedition.

9.2.1.2. Project SpeCrep: speciation processes in butterflies

Participants: Dominique Lavenier, Jeremy Gauthier, Fabrice Legeai, Claire Lemaitre, Pierre Peterlongo.

Coordinator: M. Elias (Museum National d'Histoire Naturelle, Institut de Systematique et d'Evolution de la Biodiversite, Paris)

Duration: 48 months (Jan. 2015 – Dec. 2018)

Partners: MNHN (Paris), INRA (Versailles-Grignon), Genscale Inria/IRISA Rennes.

The SpeCrep project aims at better understanding the speciation processes, in particular by comparing natural replicates from several butterfly species in a suture zone system. GenScale's task is to develop new efficient methods for the assembly of reference genomes and the evaluation of the genetic diversity in several butterfly populations.

# 9.2.2. PIA: Programme Investissement d'Avenir

9.2.2.1. RAPSODYN: Optimization of the rapeseed oil content under low nitrogen Participants: Dominique Lavenier, Claire Lemaitre, Sebastien Letort, Pierre Peterlongo.

Coordinator: N. Nesi (Inra, IGEPP, Rennes) Duration:7.5 year (2012-2019) Partners: 5 companies, 9 academic research labs. The objective of the Rapsodyn project is the optimization of the rapeseed oil content and yield under low nitrogen input. GenScale is involved in the bioinformatics work package to elaborate advanced tools dedicated to polymorphism and application to the rapeseed plant. (http://www.rapsodyn.fr)

#### 9.2.2.2. Institut Français de Bioinformatique: Plant node

Participant: Fabrice Legeai.

Coordinator: Hadi Quesneville (INRA, Versailles)

The aim of the Institut Francais de Bioinformatique (IFB) offers resources for a large community of French biologist. With INRA and CIRAD, we were part of the plant node of IFB, and focused on delivering efficient tools for sharing agronomical data, such as Askomics.

# 9.2.3. Programs from research institutions

9.2.3.1. Inria ADT DiagCancer

Participants: Dominique Lavenier, Patrick Durand.

Since October 1st, 2016, Genscale started a one-year Inria ADT called DiagCancer. It aims at: (1) including the DiscoSnp++ tool within the current data production pipeline at Pontchaillou Hospital (Rennes), (2) providing a new prediction tool applied to the calling of cancer related mutations from DNA sequencing data and (3) creating new analysis tools to facilitate the interpretation of results by end-users (biologists, doctors). The project is done in close collaboration with Haematology Service, CHU Pontchaillou, Rennes.

#### 9.2.3.2. CNRS Mastodons program: C3G

Participants: Dominique Lavenier, Pierre Peterlongo, Claire Lemaitre, Camille Marchet, Lolita Lecompte.

High-throughput sequencing applications now cover all life sciences: from medicine to agronomy. The 3rd generation sequencing produces very long reads, but the reads are extremely noisy, which has a strong impact on the quality of bioinformatics analyses. The challenge of the C3G project is to bring this type of data to a high level of quality through the development of new correction strategies.

#### 9.2.3.3. Inria Project Lab: Neuromarkers

Participants: Dominique Lavenier, Pierre Peterlongo, Claire Lemaitre.

The IPL Neuromarkers aims to design imaging biomarkers of neurodegenerative diseases for clinical trials and study of their genetic associations. In this project, GenScale bring its expertise in the genomic field.

# 9.3. International Initiatives

# 9.3.1. Inria Associate Team: HipcoGen

- Title: High-Performance Combinatorial Optimization for Computational Genomics
- International Partner (Institution Laboratory Researcher):
  - Los Alamos National Laboratory (LANL)-NM, United States, CCS-3, Hristo Djidjev
- Start year: 2017
- Teams' web site: https://team.inria.fr/genscale/presentation/associated-team/

Genome sequencing and assembly, the determination of the DNA sequences of a genome, is a core experiment in computational biology. During the last decade, the cost of sequencing has decreased dramatically and a huge amount of new genomes have been sequenced. Nevertheless, most of recent genome projects stay unfinished and nowadays the databases contain much more incompletely assembled genomes than whole stable reference genomes. The main reason is that producing a complete genome, or an as-complete-as-possible-genome, is an extremely difficult computational task (an NP-hard problem) and, in spite of the efforts and the progress done by the bioinformatics community, no satisfactory solution is available today. New sequencing technologies (such as PacBio or Oxford Nanopore) are being developed that tend to produce longer DNA sequences and offer new opportunities, but also bring significant new challenges. The goal of this joint project–a cooperation between Los Alamos National Laboratory, US and Inria, is to develop a new methodology and tools based on novel optimization techniques and massive parallelism suited to these emerging technologies and able to tackle the complete assembly of large genomes.

# 9.3.2. Informal International Partners

• Free University of Brussels, Belgium: Genome assembly [P. Perterlongo, A. Limasset]

# 9.4. International Research Visitors

# 9.4.1. Visits of International Scientists

• Visit of Hristo Djidjev from Los Alamos National Laboratory, June 5 to July 4, 2017

# 9.4.2. Visits to International Teams

- Visit of R. Andonov at LANL from May 4th to May 30th. Work on Task 2 from HipcoGen project.
- Visit of S. Francois at LANL from May 4th to May 30th and from August 2 to August 23. Work on Task 2 from HipcoGen project.
- Visit of Pierre Peterlongo at LANL, May 2017 (one week). Talk to SFAF conference: "Assembly of heterozygous genomes".

# **10.** Dissemination

# **10.1. Promoting Scientific Activities**

# 10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

• RCAM 2017: Workshop "Recent Computational Advances in Metagenomics" [P. Peterlongo]

# 10.1.2. Scientific Events Selection

# 10.1.2.1. Member of the Conference Program Committees

- JOBIM 2017: French symposium of Bioinformatics [C. Lemaitre, P. Peterlongo]
- BIBM 2017: IEEE International Conference on Bioinformatics and Biomedicine [D. Lavenier]
- BIOKDD 2017: Workshop on Biological Knowledge Discovery and Data Mining [D. Lavenier]
- IWBBIO 2017: International Work-Conference on Bioinformatics and Biomedical Engineering [D. Lavenier]

10.1.2.2. Reviewer

- RECOMB 2017 [R. Andonov, C. Lemaitre, P. Peterlongo]
- ECCB 2017 [P. Peterlongo]

# 10.1.3. Journal

10.1.3.1. Reviewer - Reviewing Activities

- Bioinformatics [D. Lavenier]
- BMC Bioinformtics [D. Lavenier, C. Marchet]
- Briefing in Bioinformatics [D. Lavenier]
- Plos One [C. Lemaitre]
- Transactions on Computational Biology and Bioinformatics [C. Lemaitre]
- Genomics [F. Legeai]

# 10.1.4. Invited Talks

- D. Lavenier, *Dealing with mass of genomic data. From optimized data structures to advanced memory architectures*, BIATA 2017, Dioinformatics: from algorithms to application, St, Petersburg, Russia, Aug. 2017
- D. Lavenier, DNA Mapping using Processing-in-Memory Architecture, Cristal, Univ. Lille, Oct 2017
- C. Lemaitre, *Looking for genomic variants in the De Bruijn Graph.*, Institute for Advanced Biosciences, University of Grenoble Alpes, Grenoble, France, Dec. 2017
- C. Marchet, *De novo Clustering of Gene Expressed Variants in Transcriptomic Long Reads Data Sets*, at From RNA-Seq data to bioinformatics analysis using Nanopore sequencers, Evry, Dec 2017.

# 10.1.5. Leadership within the Scientific Community

- P. Peterlongo. Animator of the metagenomic scientific axes of the GDR BIM (National Researc Group in Biology, Informatic and Mathematic)
- P. Peterlongo. Member of the SFBI board.

# 10.1.6. Scientific Expertise

- Expert for the MEI (International Expertise Mission), French Research Ministry [D. Lavenier]
- Member of the Scientific Council of BioGenOuest [D. Lavenier]
- Member of the Scientific Council of the Computational Biology Institute of Montpellier [D. Lavenier]

# 10.1.7. Research Administration

- Member of the CoNRS, section 06, [D. Lavenier]
- Member of the CoNRS, section 51, [D. Lavenier]
- Member of the steering committee of the INRA BIPAA Platform (BioInformatics Platform for Agroecosystems Arthropods) [D. Lavenier]
- Member of the steering committee of The GenOuest Platform (Bioinformatics Platform of BioGenOuest) [D. Lavenier]
- Representative of the environnemental axis of UMR IRISA [C. Lemaitre]
- AGOS first secretary [P. Peterlongo]
- Organisation of the weekly seminar "Symbiose" [P. Peterlongo]
- In charge of the bachelor's degree in the computer science department of University of Rennes 1 (90 students) [R. Andonov]
- Member of the Council of Administration of ISTIC [R. Andonov]
- Representative of non-permanent members in the Inria Rennes center commitee [S. Letort]

# 10.2. Teaching - Supervision - Juries

# 10.2.1. Teaching

Licence : C. Lemaitre, Statistics for biology, 30h, L3, Univ. Rennes 1, France.

Licence : R. Andonov, S. Francois, Graph Algorithms, 80h, L3, Univ. Rennes 1, France.

Licence : L. Lecompte, Systems, 16h, L3, Univ. Rennes 1, France.

Licence : S. Francois, Programming, 22h, L3 Miage, Univ. Rennes 1, France.

Master : R. Andonov, S. Francois, Operational research, 82h, M1 Miage, Univ. Rennes 1, France.

Master : L. Lecompte, Python for ecologists, 21h, M1, Univ. Rennes 1, France.

Master : C. Lemaitre, P. Peterlongo, Algorithms on Sequences, 52h, M2, Univ. Rennes 1, France.

Master : L. Lecompte, C. Lemaitre, P. Peterlongo, Algorithms on Sequences for Bioinformatics, 50h, M1, Univ. Rennes 1, France.

Master : C. Lemaitre, P. Peterlongo, Experimental Bioinformactics, 24h, M1, ENS Rennes, France. Master : C. Guyomar, Statistical learnings, 30h, M2, Univ. Rennes, France.

Master : F. Legeai, RNA-Seq, Metagenomics and Variant discovery, 12h, M2, AgroCampusOuest, National Superior School Of Agronomy, Rennes, France.

Master : R. Andonov, Advanced Algorithmics, 25h, Univ. Rennes 1, France.

Training : P. Durand, G. Rizk, GATB Programming Day, 16h (April 26-27), Grenada, Spain.

Training : P. Durand, G. Rizk, GATB Programming Day, 8h (June 6), Montpellier, France.

Training : G. Rizk, GATB Programming Day, 8h (October 24), Rennes, France.

# 10.2.2. Supervision

PhD : G. Benoit, Large scale de novo comparative metagenomics, Univ Rennes, 29/11/2017, D. Lavenier and C. Lemaitre. [11]

PhD : A. Limasset, Nouvelles approches pour l'exploitation des données de séquençage haut débit, Univ Rennes, 12/07/2017, D. Lavenier and P. Peterlongo. [12]

PhD : P. Hoan Son, Novel Pattern Mining Techniques for Genome-wide Association Studies, 22/12/2017, D. Lavenier and A. Termier.

PhD in progress : C. Guyomar, Bioinformatic tools and applications for metagenomics of bacterial communities associated to insects, 01/10/2015, J.C. Simon, C. Mougel, C. Lemaitre and F. Legeai.

PhD in progress : C. Marchet, Nouvelles méthodologies pour l'assemblage de données de séquençage polymorphes, 01/10/2015, P. Peterlongo.

PhD in progress : S. François, Combinatorial Optimization Approaches for Bioinformatics, 01/10/2016, R. Andonov.

PhD in progress : L. Lecompte, Structural Variant detection in long-read sequencing data, 01/09/2017, D. Lavenier and C. Lemaitre.

PhD in progress : W. Delage, Assemblage de novo local pour la détection de variations complexes pour le diagnostic des maladies rares, 01/10/2017, J. Thévenon and C. Lemaitre.

# 10.2.3. Juries

- *Member of Ph-D thesis juries*. Arnaud Meng, University Pierre et Marie Curie [C. Lemaitre], Damien Courtine, University of Brest [D. Lavenier].
- *Referee of Ph-D thesis.* Pierre Pericard, University of Lille [D. Lavenier], Louise-Amelie Schmitt, Bordeaux University [P. Peterlongo], Kamil Salikov, University Paris Est [P. Peterlongo]
- Member of Ph-D thesis comitees. L. Ishi Soares de Lima, University of Lyon [C. Lemaitre], Cervin Guyomar, University of Rennes [P. Peterlongo], Pierre Marijon, University of Lille [P. Peterlongo], Pierre Charrier, Oniris Nantes [P. Peterlongo], Victor Gaborit, Inserm Nantes [P. Peterlongo], Guillaume Gautreau, University Paris Saclay [P. Peterlongo].
- President of a jury for the recuitment of a INRA bioinformatics engineer [Fabrice Legeai].

# **10.3.** Popularization

- Participation to operation "A la découverte de la recherche" in high schools [P. Peterlongo]
- Intervention for "Imagine For Margo" [P. Peterlongo]
- Bioinformatics introduction to secondary school pupils [F. Legeai]

# 11. Bibliography

# Major publications by the team in recent years

 R. ANDONOV, N. MALOD-DOGNIN, N. YANEV. Maximum Contact Map Overlap Revisited, in "Journal of Computational Biology", January 2011, vol. 18, n<sup>o</sup> 1, p. 1-15 [DOI: 10.1089/CMB.2009.0196], http://hal. inria.fr/inria-00536624/en.

- [2] G. BENOIT, P. PETERLONGO, M. MARIADASSOU, E. DREZEN, S. SCHBATH, D. LAVENIER, C. LEMAITRE. Multiple comparative metagenomics using multiset k -mer counting, in "PeerJ Computer Science", November 2016, vol. 2 [DOI: 10.7717/PEERJ-CS.94], https://hal.inria.fr/hal-01397150.
- [3] R. CHIKHI, G. RIZK.Space-efficient and exact de Bruijn graph representation based on a Bloom filter, in "Algorithms for Molecular Biology", 2013, vol. 8, n<sup>o</sup> 1, 22 [DOI: 10.1186/1748-7188-8-22], http://hal. inria.fr/hal-00868805.
- [4] E. DREZEN, G. RIZK, R. CHIKHI, C. DELTEL, C. LEMAITRE, P. PETERLONGO, D. LAVE-NIER.GATB: Genome Assembly & Analysis Tool Box, in "Bioinformatics", 2014, vol. 30, p. 2959 -2961 [DOI: 10.1093/BIOINFORMATICS/BTU406], https://hal.archives-ouvertes.fr/hal-01088571.
- [5] N. MAILLET, C. LEMAITRE, R. CHIKHI, D. LAVENIER, P. PETERLONGO. Compareads: comparing huge metagenomic experiments, in "RECOMB Comparative Genomics 2012", Niterói, Brazil, October 2012, https:// hal.inria.fr/hal-00720951.
- [6] N. MALOD-DOGNIN, R. ANDONOV, N. YANEV. Maximum Cliques in Protein Structure Comparison, in "SEA 2010 9th International Symposium on Experimental Algorithms", Naples, Italy, P. FESTA (editor), Springer, May 2010, vol. 6049, p. 106-117 [DOI : 10.1007/978-3-642-13193-6\_10], https://hal.inria.fr/ inria-00536700.
- [7] V. H. NGUYEN, D. LAVENIER. PLAST: parallel local alignment search tool for database comparison, in "Bmc Bioinformatics", October 2009, vol. 10, 24, http://hal.inria.fr/inria-00425301.
- [8] G. RIZK, A. GOUIN, R. CHIKHI, C. LEMAITRE.*MindTheGap: integrated detection and assembly of short and long insertions*, in "Bioinformatics", December 2014, vol. 30, n<sup>o</sup> 24, p. 3451 3457 [DOI: 10.1093/BIOINFORMATICS/BTU545], https://hal.inria.fr/hal-01081089.
- [9] G. RIZK, D. LAVENIER.GASSST: Global Alignment Short Sequence Search Tool, in "Bioinformatics", August 2010, vol. 26, n<sup>o</sup> 20, p. 2534-2540, http://hal.archives-ouvertes.fr/hal-00531499.
- [10] R. URICARU, G. RIZK, V. LACROIX, E. QUILLERY, O. PLANTARD, R. CHIKHI, C. LEMAITRE, P. PETERLONGO.*Reference-free detection of isolated SNPs*, in "Nucleic Acids Research", November 2014, p. 1 12 [DOI: 10.1093/NAR/GKU1187], https://hal.inria.fr/hal-01083715.

# **Publications of the year**

# **Doctoral Dissertations and Habilitation Theses**

- [11] G. BENOIT.Large-scale de novo comparative metagenomics, Université Rennes1, November 2017, https:// hal.inria.fr/tel-01659395.
- [12] A. LIMASSET. Novel approaches for the exploitation of high throughput sequencing data, Université Rennes 1, July 2017, https://hal.archives-ouvertes.fr/tel-01566938.
- [13] H. S. PHAM.Novel Pattern Mining Techniques for Genome-wide Association Studies, IRISA, equipe GENSCALE, December 2017, https://hal.inria.fr/tel-01672442.

# **Articles in International Peer-Reviewed Journal**

- [14] D. EOCHE-BOSY, M. GAUTIER, M. ESQUIBET, F. LEGEAI, A. BRETAUDEAU, O. BOUCHEZ, S. FOURNET, E. GRENIER, J. MONTARRY. Genome scans on experimentally evolved populations reveal candidate regions for adaptation to plant resistance in the potato cyst nematode Globodera pallida, in "Molecular Ecology", July 2017, vol. 26, n<sup>o</sup> 18, p. 4700–4711 [DOI : 10.1111/MEC.14240], https://hal.inria.fr/hal-01605681.
- [15] A. GOUIN, A. BRETAUDEAU, K. NAM, S. GIMENEZ, J.-M. AURY, B. DUVIC, F. HILLIOU, N. DURAND, N. MONTAGNÉ, I. DARBOUX, S. KUWAR, T. CHERTEMPS, D. SIAUSSAT, A. BRETSCHNEIDER, Y. MONÉ, S.-J. AHN, S. HÄNNIGER, A.-S. GOSSELIN GRENET, D. NEUNEMANN, F. MAUMUS, I. LUYTEN, K. LABADIE, W. XU, F. KOUTROUMPA, J.-M. ESCOUBAS, A. LLOPIS, M. MAÏBÈCHE-COISNE, F. SALASC, A. TOMAR, A. R. ANDERSON, S. A. KHAN, P. DUMAS, M. ORSUCCI, J. GUY, C. BELSER, A. ALBERTI, B. NOEL, A. COULOUX, J. MERCIER, S. NIDELET, E. DUBOIS, N.-Y. LIU, I. BOULOGNE, O. MIRABEAU, G. LE GOFF, K. GORDON, J. OAKESHOTT, F. L. CONSOLI, A.-N. VOLKOFF, H. W. FESCEMYER, J. H. MARDEN, D. S. LUTHE, S. HERRERO, D. G. HECKEL, P. WINCKER, G. J. KERGOAT, J. AMSELEM, H. QUESNEVILLE, A. T. GROOT, E. JACQUIN-JOLY, N. NÈGRE, C. LEMAITRE, F. LEGEAI, E. D'ALENÇON, P. FOURNIER. Two genomes of highly polyphagous lepidopteran pests (Spodoptera frugiperda, Noctuidae) with different host-plant ranges, in "Scientific Reports", December 2017, vol. 7, n<sup>0</sup> 1, p. 1-12 [DOI: 10.1038/s41598-017-10461-4], https://hal.inria.fr/hal-01633879.
- [16] L. LIMA, B. SINAIMERI, G. SACOMOTO, H. LOPEZ-MAESTRE, C. MARCHET, V. MIELE, M.-F. SAGOT, V. LACROIX. Playing hide and seek with repeats in local and global de novo transcriptome assembly of short RNA-seq reads, in "Algorithms for Molecular Biology", December 2017, vol. 12, n<sup>o</sup> 1, 2 [DOI: 10.1186/s13015-017-0091-2], https://hal.inria.fr/hal-01474524.
- [17] T. C. MATHERS, Y. CHEN, G. KAITHAKOTTIL, F. LEGEAI, S. T. MUGFORD, P. BAA-PUYOULET, A. BRETAUDEAU, B. CLAVIJO, S. COLELLA, O. COLLIN, T. DALMAY, T. DERRIEN, H. FENG, T. GABALDON, A. JORDAN, I. JULCA, G. J. KETTLES, K. KOWITWANICH, D. LAVENIER, P. LENZI, S. LOPEZ-GOMOLLON, D. LOSKA, D. MAPLESON, F. MAUMUS, S. MOXON, D. R. G. PRICE, A. SUGIO, M. V. MUNSTER, M. UZEST, D. WAITE, G. JANDER, D. TAGU, A. C. C. WILSON, C. VAN OOSTERHOUT, D. SWARBRECK, S. A. HOGENHOUT. *Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species*, in "Genome Biology", 2017, vol. 18, n<sup>o</sup> 1, 27 [DOI: 10.1186/s13059-016-1145-3], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01500475.
- [18] G. RICHARD, F. LEGEAI, N. PRUNIER-LETERME, A. BRETAUDEAU, D. TAGU, J. JAQUIÉRY, G. LE TRIONNAIRE. Dosage compensation and sex-specific epigenetic landscape of the X chromosome in the pea aphid, in "Epigenetics & Chromatin", June 2017, vol. 10, n<sup>o</sup> 1, 30 [DOI: 10.1186/s13072-017-0137-1], https://hal.inria.fr/hal-01555242.
- [19] A. ROCHEFORT, S. BOUKTHIR, S. MOULLEC, A. MEYGRET, Y. ADNANI, D. LAVENIER, A. FAILI, S. KAYAL.Full Sequencing and Genomic Analysis of Three emm75 Group A Streptococcus Strains Recovered in the Course of an Epidemiological Shift in French Brittany, in "Genome Announcements", 2017, vol. 5, n<sup>o</sup> 39, e00957 [DOI: 10.1128/GENOMEA.00957-17], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01617890.
- [20] A. SCZYRBA, P. HOFMANN, P. BELMANN, D. KOSLICKI, S. JANSSEN, J. DRÖGE, I. GREGOR, S. MAJDA, J. FIEDLER, E. DAHMS, A. BREMGES, A. FRITZ, R. GARRIDO-OTER, T. S. JØRGENSEN, N. SHAPIRO, P. D. BLOOD, A. GUREVICH, Y. BAI, D. TURAEV, M. Z. DEMAERE, R. CHIKHI, N. NAGARAJAN, C. QUINCE, L. H. HANSEN, S. J. SØRENSEN, B. K. H. CHIA, B. DENIS, J. L. FROULA, Z. WANG, R. EGAN, D. DON KANG, J. J. COOK, C. DELTEL, M. BECKSTETTE, C. LEMAITRE, P. PETERLONGO, G. RIZK, D. LAVENIER, Y.-W. WU, S. W. SINGER, C. JAIN, M. STROUS, H. KLINGENBERG, P. MEINICKE,

M. D. BARTON, T. LINGNER, H.-H. LIN, Y.-C. LIAO, G. G. Z. SILVA, D. A. CUEVAS, R. A. EDWARDS, S. SAHA, V. C. PIRO, B. Y. RENARD, M. POP, H.-P. KLENK, M. GÖKER, N. C. KYRPIDES, T. WOYKE, J. A. VORHOLT, P. SCHULZE-LEFERT, E. M. RUBIN, A. E. DARLING, T. RATTEI, A. C. MCHARDY.*Critical Assessment of Metagenome Interpretation – a benchmark of computational metagenomics software*, in "Nature Methods", October 2017, vol. 14, n<sup>0</sup> 11, p. 1063 - 1071 [*DOI* : 10.1038/NMETH.4458], https://hal.archives-ouvertes.fr/hal-01633525.

- [21] S. C. VARMA, P. KOLIN, M. BALAKRISHNAN, D. LAVENIER. Hardware acceleration of de novo genome assembly, in "International Journal of Embedded Systems", February 2017, vol. 9, n<sup>o</sup> 1, p. 74-89 [DOI: 10.1504/IJES.2017.10002593], https://hal.inria.fr/hal-01481800.
- [22] J. A. WENGER, B. J. CASSONE, F. LEGEAI, J. S. JOHNSTON, R. BANSAL, A. D. YATES, B. S. COATES, V. A. C. PAVINATO, A. MICHEL. Whole genome sequence of the soybean aphid, Aphis glycines, in "Insect Biochemistry and Molecular Biology", January 2017, vol. 18, n<sup>o</sup> 1, 27 [DOI: 10.1016/J.IBMB.2017.01.005], https://hal.inria.fr/hal-01555244.
- [23] V. WUCHER, F. LEGEAI, B. HEDAN, G. RIZK, L. LAGOUTTE, T. LEEB, V. JAGANNATHAN, E. CADIEU, A. DAVID, H. LOHI, S. CIRERA, M. FREDHOLM, N. BOTHEREL, P. A. J. LEEGWATER, C. LE BEGUEC, H. FIETEN, J. JOHNSON, J. ALFÖLDI, C. ANDRÉ, K. LINDBLAD-TOH, C. HITTE, T. DERRIEN. *FEELnc: a* tool for long non-coding RNA annotation and its application to the dog transcriptome, in "Nucleic Acids Research", 2017, vol. 45, n<sup>o</sup> 8, 12 [DOI : 10.1093/NAR/GKW1306], https://hal-univ-rennes1.archivesouvertes.fr/hal-01532061.

# **Invited Conferences**

[24] V. WUCHER, F. LEGEAI, L. BOURNEUF, T. DERRIEN, A. GALLOT, S. HUDAVERDIAN, S. JAUBERT-POSSAMAI, N. LETERME-PRUNIER, J. NICOLAS, H. SEITZ, A. SIEGEL, S. TANGUY, G. LE TRION-NAIRE, D. TAGU. *Integrative genomics and gene networks for studying phenotypic plasticity in the pea aphid*, in "10th Arthropod Genomics Symposium", Notre Dame, United States, June 2017, https://hal.inria.fr/hal-01566438.

# **International Conferences with Proceedings**

- [25] S. FRANCOIS, R. ANDONOV, D. LAVENIER, H. DJIDJEV. Global optimization approach for circular and chloroplast genome assembly, in "BICoB 2018 10th International Conference on Bioinformatics and Computational Biology", Las Vegas, United States, March 2018, https://hal.inria.fr/hal-01666830.
- [26] S. FRANÇOIS, R. ANDONOV, D. LAVENIER, H. DJIDJEV. Global Optimization for Scaffolding and Completing Genome Assemblies, in "International Network Optimization Conference", Lisboa, Portugal, European Network Optimization Group (ENOG), February 2017, 15, https://hal.inria.fr/hal-01499859.
- [27] A. LIMASSET, G. RIZK, R. CHIKHI, P. PETERLONGO. Fast and scalable minimal perfect hashing for massive key sets, in "16th International Symposium on Experimental Algorithms", London, United Kingdom, June 2017, vol. 11, p. 1 - 11, https://arxiv.org/abs/1702.03154, https://hal.inria.fr/hal-01566246.

# **Conferences without Proceedings**

[28] X. GARNIER, A. BRETAUDEAU, O. FILANGI, F. LEGEAI, A. SIEGEL, O. DAMERON. AskOmics, a web tool to integrate and query biological data using semantic web technologies, in "JOBIM 2017 - Journées Ouvertes en Biologie, Informatique et Mathématiques", Lille, France, July 2017, 1, https://hal.inria.fr/hal-01577425.

- [29] X. GARNIER, O. DAMERON, O. FILANGI, F. LEGEAI, A. BRETAUDEAU. Integration of Linked Data into Galaxy using Askomics, in "Galaxy Community Conference", Montpellier, France, June 2017, https://hal. inria.fr/hal-01576870.
- [30] C. GUYOMAR, F. LEGEAI, C. MOUGEL, C. LEMAITRE, J.-C. SIMON. Multi-scale characterization of symbiont diversity in the pea aphid complex through metagenomic approaches, in "International Conference on Holobionts", Paris, France, April 2017, https://hal.archives-ouvertes.fr/hal-01638839.
- [31] A. MENG, E. CORRE, P. PETERLONGO, C. MARCHET, A. ALBERTI, C. D. SILVA, P. WINCKER, I. PROBERT, N. SUZUKI, S. LE CROM, L. BITTNER, F. NOT. A transcriptomic approach to study marine plankton holobionts, in "International Conference on Holobionts", Paris, France, April 2017, https://hal.inria. fr/hal-01575069.
- [32] H.-S. PHAM, J.-H. FERRASSE, O. BOUTIN, N. ALPY, M. SAEZ. Analyse thermodynamique du cycle de conversion utilisant le CO2 Supercritique pour une application en réacteur modulaire, in "SFGP 2017", Nancy, France, July 2017, https://hal.archives-ouvertes.fr/hal-01679087.

# Scientific Books (or Scientific Book chapters)

[33] G. BENOIT, C. LEMAITRE, G. RIZK, E. DREZEN, D. LAVENIER. *De Novo NGS Data Compression*, in "Algorithms for Next-Generation Sequencing Data", M. ELLOUMI (editor), Springer, 2017, p. 91-115 [*DOI*: 10.1007/978-3-319-59826-0\_4], https://hal.archives-ouvertes.fr/hal-01633718.

#### **Research Reports**

- [34] L. BOURI, D. LAVENIER. Evaluation of long read error correction software, Inria Rennes Bretagne Atlantique; GenScale, February 2017, n<sup>o</sup> RR-9028, https://hal.inria.fr/hal-01463694.
- [35] L. BOURI, D. LAVENIER, J.-F. J.-F. GIBRAT, V. F. DOMINGUEZ DEL ANGEL. Evaluation of genome assembly software based on long reads, France Genomique, March 2017 [DOI: 10.5281/ZENODO.345098], https://hal.inria.fr/hal-01481801.

# **Other Publications**

- [36] G. BENOIT, P. PETERLONGO, M. MARIADASSOU, E. DREZEN, S. SCHBATH, D. LAVENIER, C. LEMAITRE.Simka: large scale de novo comparative metagenomics, July 2017, 234, JOBIM 2017 Journées Ouvertes Biologie Informatique Mathématiques, Poster Acte onlineReference paper: Benoit et al. (2016) Multiple comparative metagenomics using multiset k-mer counting. PeerJComputer Science. https://doi.org/10.7717/peerj-cs.94, https://hal.archives-ouvertes.fr/hal-01595071.
- [37] C. BENOIT-PILVEN, C. MARCHET, E. CHAUTARD, L. LIMA, M.-P. LAMBERT, G. SACOMOTO, A. REY, C. BOURGEOIS, D. AUBOEUF, V. LACROIX. Annotation and differential analysis of alternative splicing using de novo assembly of RNAseq data, November 2017, working paper or preprint [DOI : 10.1101/074807], https://hal.archives-ouvertes.fr/hal-01643169.
- [38] C. DIOT, F. HERAULT, J. NAVARRO, L. LE CALVEZ, E. BAEZA, C. KLOPP, O. BOUCHEZ, D. ESQUERRE, P. PETERLONGO. *Genome specific expression in the liver of mule and hinny duck hybrids*, June 2017, 1, 10. European Symposium on Poultry Genetics (ESPG), Poster, https://hal.archives-ouvertes.fr/hal-01594559.
- [39] J. GAUTHIER, C. MOUDEN, T. SUCHAN, N. ALVAREZ, N. ARRIGO, C. RIOU, C. LEMAITRE, P. PE-TERLONGO. *DiscoSnp-RAD: de novo detection of small variants for population genomics*, November 2017, working paper or preprint, https://hal.inria.fr/hal-01634232.
- [40] C. GUYOMAR, F. LEGEAI, C. MOUGEL, C. LEMAITRE, J.-C. SIMON.Multi-scale characterization of symbiont diversity in the pea aphid complex through metagenomic approaches, July 2017, JOBIM 2017 -Journées Ouvertes en Biologie, Informatique et Mathématiques, Poster, https://hal.archives-ouvertes.fr/hal-01638884.
- [41] A. LIMASSET, J.-F. FLOT, P. PETERLONGO. Toward perfect reads, November 2017, https://arxiv.org/abs/ 1711.03336 - RECOMB 2018 Submission, https://hal.inria.fr/hal-01644163.
- [42] C. MARCHET, L. LECOMPTE, C. DA SILVA, C. CRUAUD, J.-M. AURY, J. NICOLAS, P. PETERLONGO.De novo Clustering of Gene Expressed Variants in Transcriptomic Long Reads Data Sets, November 2017, working paper or preprint [DOI: 10.1101/170035], https://hal.archives-ouvertes.fr/hal-01643156.
- [43] C. MARCHET, L. LECOMPTE, A. LIMASSET, L. BITTNER, P. PETERLONGO. A resource-frugal probabilistic dictionary and applications in bioinformatics, November 2017, https://arxiv.org/abs/1703.00667 - Submitted to Journal of Discrete Algorithms. arXiv admin note: substantial text overlap with arXiv:1605.08319, https:// hal.archives-ouvertes.fr/hal-01643162.
- [44] A. MENG, C. MARCHET, E. CORRE, P. PETERLONGO, A. ALBERTI, C. DA SILVA, P. WINCKER, E. PELLETIER, I. PROBERT, J. DECELLE, S. LE CROM, F. NOT, L. BITTNER. A de novo approach to disentangle partner identity and function in holobiont systems, November 2017, working paper or preprint [DOI: 10.1101/221424], https://hal.archives-ouvertes.fr/hal-01643153.

#### **References in notes**

[45] J. T. SIMPSON, K. WONG, S. D. JACKMAN, J. E. SCHEIN, S. J. M. JONES, I. BIROL*ABySS: a parallel assembler for short read sequence data*, in "Genome Res", Jun 2009, vol. 19, n<sup>0</sup> 6, p. 1117–1123.

## **Project-Team HYBRID**

# 3D interaction with virtual environments using body and mind

IN PARTNERSHIP WITH: Institut national des sciences appliquées de Rennes

**Université Rennes 1** 

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Interaction and visualization

## **Table of contents**

1.	Personnel	366
2.	Overall Objectives	367
3.	Research Program	368
	3.1. Research Program	368
	3.2. Research Axes	369
4.	Application Domains	. 370
5.	Highlights of the Year	370
6.	New Software and Platforms	371
	6.1. #FIVE	371
	6.2. #SEVEN	371
	6.3. OpenVIBE	372
_	6.4. Platforms	372
7.	New Results	373
	7.1. Virtual Reality Tools and Usages	373
	7.1.1. Gesture recognition for VR	373
	7.1.2. Automatic tools for the evaluation of VR systems	373
	7.1.3. Customer behavior and analyses in VR	374
	7.2. Physically-Based Simulation and Haptic Feedback	375
	7.2.1. Physically-based simulation	375
	7.2.2. Haptic feedback	3//
	7.3. Augmented Reality	311
	7.3.1. Perception in augmented reality	270
	7.3.2. Interaction in augmented reality	380
	7.5.5. Hacking	200
	7.4. Brann-Computer Interfaces	201
	7.4.1. Definitious and techniques	382
	7.5 Cultural Heritage	385
	7.5.1 VR and AR tools for cultural heritage	385
	7.5.2 Multi-modal images and 3D printing for cultural heritage	387
	7.5.3. Generating 3D data for cultural heritage	390
8.	Bilateral Contracts and Grants with Industry	
	8.1. Bilateral Contracts with Industry	391
	8.2. Bilateral Grants with Industry	392
	8.2.1. Technicolor	392
	8.2.2. Realyz	392
	8.2.3. VINCI Construction	392
	8.2.4. Orange Labs	392
9.	Partnerships and Cooperations	393
	9.1. Regional Initiatives	393
	9.1.1. Labex Cominlabs SUNSET	393
	9.1.2. Labex Cominlabs S3PM	393
	9.1.3. Labex Cominlabs HEMISFER	393
	9.1.4. Labex Cominlabs SABRE	393
	9.1.5. IRT b<>com	394
	9.1.6. CNPAO Project	394
	9.1.7. Imag'In CNRS IRMA	394
	9.2. National Initiatives	394
	9.2.1. ANR-FRQSC INTROSPECT	394

	9.2.2. Ilab CertiViBE	395
	9.2.3. IPL BCI-LIFT	395
	9.2.4. ATT CONSORVIBE	395
	9.3. European Initiatives	395
	9.3.1.1. HAPPINESS	395
	9.3.1.2. IMAGINE	396
	9.4. International Research Visitors	397
	9.4.1. Visits of International Scientists	397
	9.4.2. Visits to International Teams	397
10.	Dissemination	
	10.1. Promoting Scientific Activities	397
	10.1.1. Scientific events organisation	397
	10.1.1.1. General chair, Scientific chair	397
	10.1.1.2. Member of the organizing committees	397
	10.1.2. Scientific events selection	397
	10.1.2.1. Member of the conference program committees	397
	10.1.2.2. Reviewer	398
	10.1.3. Journal	398
	10.1.3.1. Member of the editorial boards	398
	10.1.3.2. Reviewer - Reviewing activities	398
	10.1.4. Invited talks	398
	10.1.5. Leadership within the scientific community	398
	10.1.6. Scientific expertise	399
	10.1.7. Research administration	399
	10.2. Teaching - Supervision - Juries	399
	10.2.1. Supervision	400
	10.2.1.1. PhD (defended)	400
	10.2.1.2. PhD (in progress)	400
	10.2.2. Juries	401
	10.2.2.1. Selection committees	401
	10.2.2.2. PhD and HDR juries	401
	10.3. Popularization	402
11.	Bibliography	402

### **Project-Team HYBRID**

*Creation of the Team: 2013 January 01, updated into Project-Team: 2013 July 01* **Keywords:** 

#### **Computer Science and Digital Science:**

A2.5. - Software engineering

A5. - Interaction, multimedia and robotics

A5.1. - Human-Computer Interaction

A5.1.2. - Evaluation of interactive systems

A5.1.3. - Haptic interfaces

A5.1.4. - Brain-computer interfaces, physiological computing

A5.1.5. - Body-based interfaces

A5.1.6. - Tangible interfaces

A5.1.7. - Multimodal interfaces

A5.1.8. - 3D User Interfaces

A5.5.4. - Animation

A5.6. - Virtual reality, augmented reality

A6. - Modeling, simulation and control

A6.2. - Scientific Computing, Numerical Analysis & Optimization

A6.3. - Computation-data interaction

#### **Other Research Topics and Application Domains:**

B1.2. - Neuroscience and cognitive science

- B2. Health
- B2.4. Therapies

B2.5. - Handicap and personal assistances

B2.6. - Biological and medical imaging

B2.7. - Medical devices

B2.7.1. - Surgical devices

B2.8. - Sports, performance, motor skills

- B5. Industry of the future
- B5.1. Factory of the future
- B5.2. Design and manufacturing
- B5.8. Learning and training

B5.9. - Industrial maintenance

- B8.1. Smart building/home
- B8.3. Urbanism and urban planning
- B9.1. Education
- B9.2. Art
- B9.2.2. Cinema, Television
- B9.2.3. Video games
- B9.3. Sports
- B9.5.6. Archeology, History

## 1. Personnel

#### **Research Scientists**

Anatole Lécuyer [Team leader, Inria, Senior Researcher, HDR] Ferran Argelaguet Sanz [Inria, Researcher]

#### **Faculty Members**

Bruno Arnaldi [INSA Rennes, Professor, HDR] Valérie Gouranton [INSA Rennes, Associate Professor] Maud Marchal [INSA Rennes, Associate Professor, HDR] Benoit Le Gouis [Univ. Rennes I, Associate Professor, from Sep 2017]

#### **External Collaborators**

Francois Lehericey [VINCI Construction] Guillaume Moreau [Ecole Centrale de Nantes, from Mar 2017] Jean-Marie Normand [Ecole Centrale de Nantes, from Mar 2017]

#### **Technical Staff**

Florian Nouviale [INSA Rennes, SED Research Engineer, 20%] Ronan Gaugne [Univ. Rennes 1, SED Research Engineer, 15%] Kevin-Yoren Gaffary [INSA Rennes, from Oct 2017] Alexandre Audinot [INSA Rennes, from Apr 2017] Guillaume Claude [INSA Rennes, until Apr 2017] Thierry Gaugry [Inria] Emeric Goga [SATT Ouest Valorisation, from Sep 2017] Vincent Goupil [SATT Ouest Valorisation, from Sep 2017] Carl-Johan Jorgensen [SATT Ouest Valorisation, from Sep 2017] Jussi Tapio Lindgren [Inria] Marsel Mano [Inria, until Sep 2017] Cédric Riou [Inria, from May 2017]

#### PhD Students

Benoit Le Gouis [INSA Rennes, until Aug 2017] Jean-Baptiste Barreau [CNRS, until Aug 2017] Guillaume Bataille [Orange Labs, from Oct 2017] Antonin Bernardin [INSA Rennes, from Sep 2017] Lorraine Perronnet [Inria, until Apr 2017] Mathis Fleury [Inria, from Nov 2017] Guillaume Cortes [Realyz] Antoine Costes [Technicolor] Xavier de Tinguy de La Girouliere [ENS Paris-Saclay, from Sep 2017] Anne-Solène Dris-Kerdreux [VINCI Construction] Rebecca Fribourg [Inria, from Sep 2017] Romain Lagneau [INSA Rennes, from Sep 2017] Gwendal Le Moulec [INSA Rennes] Flavien Lecuyer [INSA Rennes, from Sep 2017] Etienne Peillard [Inria/ECN, from Oct 2017] Gautier Picard [INSA Rennes, until Nov 2017] Adrien Reuzeau [INSA Rennes, from Oct 2017] Hakim Si Mohammed [Inria] Romain Terrier [IRT b<>com, from Oct 2017]

#### **Post-Doctoral Fellows**

Kevin-Yoren Gaffary [Inria, until Sep 2017] Camille Jeunet [Inria, from Feb 2017] Giulia Lioi [Inria, from Nov 2017] Nataliya Kos'Myna [Inria, until May 2017]

#### **Visiting Scientist**

Théophane Nicolas [CNRS, until Jun 2017]

Administrative Assistant Nathalie Denis [Inria]

## 2. Overall Objectives

#### 2.1. Overall Objectives

Our research project belongs to the scientific field of Virtual Reality (VR) and 3D interaction with virtual environments. VR systems can be used in numerous applications such as for industry (virtual prototyping, assembly or maintenance operations, data visualization), entertainment (video games, theme parks), arts and design (interactive sketching or sculpture, CAD, architectural mock-ups), education and science (physical simulations, virtual classrooms), or medicine (surgical training, rehabilitation systems). A major change that we foresee in the next decade concerning the field of Virtual Reality relates to the emergence of new paradigms of interaction (input/output) with Virtual Environments (VE).

As for today, the most common way to interact with 3D content still remains by measuring user's motor activity, i.e., his/her gestures and physical motions when manipulating different kinds of input device. However, a recent trend consists in soliciting more movements and more physical engagement of the body of the user. We can notably stress the emergence of bimanual interaction, natural walking interfaces, and whole-body involvement. These new interaction schemes bring a new level of complexity in terms of generic physical simulation of potential interactions between the virtual body and the virtual surrounding, and a challenging "trade-off" between performance and realism. Moreover, research is also needed to characterize the influence of these new sensory cues on the resulting feelings of "presence" and immersion of the user.

Besides, a novel kind of user input has recently appeared in the field of virtual reality: the user's mental activity, which can be measured by means of a "Brain-Computer Interface" (BCI). Brain-Computer Interfaces are communication systems which measure user's electrical cerebral activity and translate it, in real-time, into an exploitable command. BCIs introduce a new way of interacting "by thought" with virtual environments. However, current BCI can only extract a small amount of mental states and hence a small number of mental commands. Thus, research is still needed here to extend the capacities of BCI, and to better exploit the few available mental states in virtual environments.

Our first motivation consists thus in designing novel "body-based" and "mind-based" controls of virtual environments and reaching, in both cases, more immersive and more efficient 3D interaction.

Furthermore, in current VR systems, motor activities and mental activities are always considered separately and exclusively. This reminds the well-known "body-mind dualism" which is at the heart of historical philosophical debates. In this context, our objective is to introduce novel "hybrid" interaction schemes in virtual reality, by considering motor and mental activities jointly, i.e., in a harmonious, complementary, and optimized way. Thus, we intend to explore novel paradigms of 3D interaction mixing body and mind inputs. Moreover, our approach becomes even more challenging when considering and connecting multiple users which implies multiple bodies and multiple brains collaborating and interacting in virtual reality.

Our second motivation consists thus in introducing a "hybrid approach" which will mix mental and motor activities of one or multiple users in virtual reality.

## 3. Research Program

#### 3.1. Research Program

The scientific objective of Hybrid team is to improve 3D interaction of one or multiple users with virtual environments, by making full use of physical engagement of the body, and by incorporating the mental states by means of brain-computer interfaces. We intend to improve each component of this framework individually, but we also want to improve the subsequent combinations of these components.

The "hybrid" 3D interaction loop between one or multiple users and a virtual environment is depicted in Figure 1. Different kinds of 3D interaction situations are distinguished (red arrows, bottom): 1) body-based interaction, 2) mind-based interaction, 3) hybrid and/or 4) collaborative interaction (with at least two users). In each case, three scientific challenges arise which correspond to the three successive steps of the 3D interaction loop (blue squares, top): 1) the 3D interaction technique, 2) the modeling and simulation of the 3D scenario, and 3) the design of appropriate sensory feedback.



Figure 1. 3D hybrid interaction loop between one or multiple users and a virtual reality system. Top (in blue) three steps of 3D interaction with a virtual environment: (1-blue) interaction technique, (2-blue) simulation of the virtual environment, (3-blue) sensory feedbacks. Bottom (in red) different cases of interaction: (1-red) body-based, (2-red) mind-based, (3-red) hybrid, and (4-red) collaborative 3D interaction.

The 3D interaction loop involves various possible inputs from the user(s) and different kinds of output (or sensory feedback) from the simulated environment. Each user can involve his/her body and mind by means of corporal and/or brain-computer interfaces. A hybrid 3D interaction technique (1) mixes mental and motor inputs and translates them into a command for the virtual environment. The real-time simulation (2) of the virtual environment is taking into account these commands to change and update the state of the virtual world and virtual objects. The state changes are sent back to the user and perceived by means of different sensory feedbacks (e.g., visual, haptic and/or auditory) (3). The sensory feedbacks are closing the 3D interaction

loop. Other users can also interact with the virtual environment using the same procedure, and can eventually "collaborate" by means of "collaborative interactive techniques" (4).

This description is stressing three major challenges which correspond to three mandatory steps when designing 3D interaction with virtual environments:

- **3D** interaction techniques: This first step consists in translating the actions or intentions of the user (inputs) into an explicit command for the virtual environment. In virtual reality, the classical tasks that require such kinds of user command were early categorized in four [33]: navigating the virtual world, selecting a virtual object, manipulating it, or controlling the application (entering text, activating options, etc). The addition of a third dimension, the use of stereoscopic rendering and the use of advanced VR interfaces make however inappropriate many techniques that proved efficient in 2D, and make it necessary to design specific interaction techniques and adapted tools. This challenge is here renewed by the various kinds of 3D interaction which are targeted. In our case, we consider various cases, with motor and/or cerebral inputs, and potentially multiple users.
- Modeling and simulation of complex 3D scenarios: This second step corresponds to the update of the state of the virtual environment, in real-time, in response to all the potential commands or actions sent by the user. The complexity of the data and phenomena involved in 3D scenarios is constantly increasing. It corresponds for instance to the multiple states of the entities present in the simulation (rigid, articulated, deformable, fluids, which can constitute both the user's virtual body and the different manipulated objects), and the multiple physical phenomena implied by natural human interactions (squeezing, breaking, melting, etc). The challenge consists here in modeling and simulating these complex 3D scenarios and meeting, at the same time, two strong constraints of virtual reality systems: performance (real-time and interactivity) and genericity (e.g., multi-resolution, multi-modal, multi-platform, etc).
- Immersive sensory feedbacks: This third step corresponds to the display of the multiple sensory feedbacks (output) coming from the various VR interfaces. These feedbacks enable the user to perceive the changes occurring in the virtual environment. They are closing the 3D interaction loop, making the user immersed, and potentially generating a subsequent feeling of presence. Among the various VR interfaces which have been developed so far we can stress two kinds of sensory feedback: visual feedback (3D stereoscopic images using projection-based systems such as CAVE systems or Head Mounted Displays); and haptic feedback (related to the sense of touch and to tactile or force-feedback devices). The Hybrid team has a strong expertize in haptic feedback, and in the design of haptic and "pseudo-haptic" rendering [34]. Note that a major trend in the community, which is strongly supported by the Hybrid team, relates to a "perception-based" approach, which aims at designing sensory feedbacks which are well in line with human perceptual capacities.

These three scientific challenges are addressed differently according to the context and the user inputs involved. We propose to consider three different contexts, which correspond to the three different research axes of the Hybrid research team, namely: 1) body-based interaction (motor input only), 2) mind-based interaction (cerebral input only), and then 3) hybrid and collaborative interaction (i.e., the mixing of body and brain inputs from one or multiple users).

#### **3.2. Research Axes**

The scientific activity of Hybrid team follows three main axes of research:

• **Body-based interaction in virtual reality.** Our first research axis concerns the design of immersive and effective "body-based" 3D interactions, i.e., relying on a physical engagement of the user's body. This trend is probably the most popular one in VR research at the moment. Most VR setups make use of tracking systems which measure specific positions or actions of the user in order to interact with a virtual environment. However, in recent years, novel options have emerged for measuring "full-body" movements or other, even less conventional, inputs (e.g. body equilibrium). In this first research axis we are thus concerned by the emergence of new kinds of "body-based interaction" with virtual environments. This implies the design of novel 3D user interfaces and novel 3D interactive

techniques, novel simulation models and techniques, and novel sensory feedbacks for body-based interaction with virtual worlds. It involves real-time physical simulation of complex interactive phenomena, and the design of corresponding haptic and pseudo-haptic feedback.

- Mind-based interaction in virtual reality. Our second research axis concerns the design of immersive and effective "mind-based" 3D interactions in Virtual Reality. Mind-based interaction with virtual environments is making use of Brain-Computer Interface technology. This technology corresponds to the direct use of brain signals to send "mental commands" to an automated system such as a robot, a prosthesis, or a virtual environment. BCI is a rapidly growing area of research and several impressive prototypes are already available. However, the emergence of such a novel user input is also calling for novel and dedicated 3D user interfaces. This implies to study the extension of the mental vocabulary available for 3D interaction with VE, then the design of specific 3D interaction techniques "driven by the mind" and, last, the design of immersive sensory feedbacks that could help improving the learning of brain control in VR.
- **Hybrid and collaborative 3D interaction.** Our third research axis intends to study the combination of motor and mental inputs in VR, for one or multiple users. This concerns the design of mixed systems, with potentially collaborative scenarios involving multiple users, and thus, multiple bodies and multiple brains sharing the same VE. This research axis therefore involves two interdependent topics: 1) collaborative virtual environments, and 2) hybrid interaction. It should end up with collaborative virtual environments with multiple users, and shared systems with body and mind inputs.

## 4. Application Domains

#### 4.1. Overview

The research program of Hybrid team aims at next generations of virtual reality and 3D user interfaces which could possibly address both the "body" and "mind" of the user. Novel interaction schemes are designed, for one or multiple users. We target better integrated systems and more compelling user experiences.

The applications of our research program correspond to the applications of virtual reality technologies which could benefit from the addition of novel body-based or mind-based interaction capabilities:

- Industry: with training systems, virtual prototyping, or scientific visualization;
- Medicine: with rehabilitation and reeducation systems, or surgical training simulators;
- Entertainment: with 3D web navigations, video games, or attractions in theme parks,
- Construction: with virtual mock-ups design and review, or historical/architectural visits.

## 5. Highlights of the Year

#### 5.1. Highlights of the Year

- The Hybrid team has considerably grown this year, reaching a total of nearly 40 team members at the end of 2017. In particular, 10 new PhD Students have been recruited in 2017, and 2 new Associate Members have joined Hybrid this year: Guillaume Moreau (Professor, Ecole Centrale de Nantes), and Jean-Marie Normand (Associate Professor, Ecole Centrale de Nantes).
- The Hybrid team was strongly involved in conference organization this year, in particular with: IEEE ISMAR 2017 and IGRV 2017. The IEEE Symposium on Mixed and Augmented Reality 2017 (IEEE ISMAR 2017) notably took place for the first time in France with around 350 attendees, in Nantes, October 9-13, with G. Moreau and A. Lécuyer: General Chairs, J.-M. Normand: Deputy General Chair, F. Argelaguet: Posters Chair, F. Nouviale: Demos Chair, V. Gouranton and R. Gaugne: VR Tour Chairs.

- The team has also organized a 1-week "hackathon" on Virtual Reality, at Inria Rennes in June 2017, with more than 20 participants and 4 teams. It was a very successful event which ended up with 4 live demos presented at Inria/IRISA Rennes and assessed by a Jury of experts.
- We have officially started to work on the topic of "Augmented Reality" this year, with a first paper published in IEEE ISMAR 2017, and several PhD students recruited on this hot topic (Etienne Peillard, Hakim Si-Mohammed, Guillaume Bataille).

#### 5.1.1. Awards

- IEEE VGTC Virtual Reality Best Dissertation Award 2017 Honorable Mention : for former PhD student Merwan Achibet for his work "Contributions to the Design of Novel Hand-based Interaction Techniques for Virtual Environments".
- GdR IG-RV Best PhD Thesis Award 2017 Honorable Mention : for former PhD student Merwan Achibet for his work "Contributions to the Design of Novel Hand-based Interaction Techniques for Virtual Environments".
- b<>com Award for Best Publication of the year 2017 for former PhD student Lucas Royer for his paper :

**BEST PAPERS AWARDS :** 

[] Medical Image Analysis. L. ROYER, A. KRUPA, G. DARDENNE, A. LE BRAS, E. MARCHAND, M. MARCHAL.

## 6. New Software and Platforms

#### 6.1. **#FIVE**

Framework for Interactive Virtual Environments

KEYWORDS: Virtual reality - 3D - 3D interaction - Behavior modeling

SCIENTIFIC DESCRIPTION: #FIVE (Framework for Interactive Virtual Environments) is a framework for the development of interactive and collaborative virtual environments. #FIVE was developed to answer the need for an easier and a faster design and development of virtual reality applications. #FIVE provides a toolkit that simplifies the declaration of possible actions and behaviours of objects in a VE. It also provides a toolkit that facilitates the setting and the management of collaborative interactions in a VE. It is compliant with a distribution of the VE on different setups. It also proposes guidelines to efficiently create a collaborative and interactive VE. The current implementation is in C# and comes with a Unity3D engine integration, compatible with MiddleVR framework.

FUNCTIONAL DESCRIPTION: #FIVE contains software modules that can be interconnected and helps in building interactive and collaborative virtual environments. The user can focus on domain-specific aspects for his/her application (industrial training, medical training, etc) thanks to #FIVE's modules. These modules can be used in a vast range of domains using virtual reality applications and requiring interactive environments and collaboration, such as in training for example.

- Participants: Florian Nouviale, Valérie Gouranton, Bruno Arnaldi, Thomas Boggini, Guillaume Claude, Thomas Lopez and Quentin Petit
- Contact: Valérie Gouranton
- Publication: #FIVE : High-Level Components for Developing Collaborative and Interactive Virtual Environments
- URL: https://bil.inria.fr/fr/software/view/2527/tab

#### **6.2. #SEVEN**

Sensor Effector Based Scenarios Model for Driving Collaborative Virtual Environments KEYWORDS: Virtual reality - Interactive Scenarios - 3D interaction

SCIENTIFIC DESCRIPTION: #SEVEN (Sensor Effector Based Scenarios Model for Driving Collaborative Virtual Environments) is a model and an engine based on petri nets extended with sensors and effectors, enabling the description and execution of complex and interactive scenarios

FUNCTIONAL DESCRIPTION: #SEVEN enables the execution of complex scenarios for driving Virtual Reality applications. #SEVEN's scenarios are based on an enhanced Petri net model which is able to describe and solve intricate event sequences. #SEVEN comes with an editor for creating, editing and remotely controlling and running scenarios. #SEVEN is implemented in C# and can be used as a stand-alone application or as a library. An integration to the Unity3D engine, compatible with MiddleVR, also exists.

- Participants: Florian Nouviale, Valérie Gouranton, Bruno Arnaldi, Guillaume Claude, Thomas Boggini and Rozenn Bouville Berthelot
- Contact: Valérie Gouranton
- Publications: Actions sequencing incollaborative virtual environment Short Paper: #SEVEN, a Sensor Effector Based Scenarios Model for Driving Collaborative Virtual Environment
- URL: https://bil.inria.fr/fr/software/view/2528/tab

#### 6.3. OpenVIBE

KEYWORDS: Neurosciences - Interaction - Virtual reality - Health - Real time - Neurofeedback - Brain-Computer Interface - EEG - 3D interaction

FUNCTIONAL DESCRIPTION: OpenViBE is a free and open-source software platform devoted to the design, test and use of Brain-Computer Interfaces (BCI). The platform consists of a set of software modules that can be integrated easily and efficiently to design BCI applications. The key features of OpenViBE software are its modularity, its high-performance, its portability, its multiple-users facilities and its connection with high-end/VR displays. The designer of the platform enables to build complete scenarios based on existing software is available on the Inria Forge under the terms of the AGPL licence, and it was officially released in June 2009. Since then, the OpenViBE software has already been downloaded more than 40000 times, and it is used by numerous laboratories, projects, or individuals worldwide. More information, downloads, tutorials, videos, documentations are available on the OpenViBE website.

- Participants: Cédric Riou, Thierry Gaugry, Anatole Lécuyer, Fabien Lotte, Jussi Tapio Lindgren, Laurent Bougrain, Maureen Clerc Gallagher and Théodore Papadopoulo
- Partners: INSERM CEA-List GIPSA-Lab
- Contact: Anatole Lécuyer
- URL: http://openvibe.inria.fr

#### 6.4. Platforms

#### 6.4.1. Immerstar

• Participants : Florian Nouviale, Ronan Gaugne

With the two platforms of virtual reality, Immersia and Immermove, grouped under the name Immerstar, the team has access to high level scientific facilities. This equipment benefits the research teams of the center and has allowed them to extend their local, national and international collaborations. The Immerstar platform is granted by a Inria CPER funding for 20152019 that enables important evolutions of the equipment. In 2017, WQXGA laser projectors were installed in Immersia as well as a new tracking system and a new cluster of computers, improving the quality, homogeneity and latency of the platform

## 7. New Results

#### 7.1. Virtual Reality Tools and Usages

#### 7.1.1. Gesture recognition for VR

#### Spatial and Rotation Invariant 3D Gesture Recognition Based on Sparse Representation

Participants: Ferran Argelaguet and Anatole Lécuyer

Advances in motion tracking technology, especially for commodity hardware, still require robust 3D gesture recognition in order to fully exploit the benefits of natural user interfaces. In this work [10], we introduced a novel 3D gesture recognition algorithm based on the sparse representation of 3D human motion. The sparse representation of human motion provides a set of features that can be used to efficiently classify gestures in real-time. Compared to existing gesture recognition systems, the proposed approach enables full spatial and rotation invariance and provides high tolerance to noise. Moreover, the proposed classification scheme takes into account the inter-user variability which increases gesture classification accuracy in user-independent scenarios. We validated our approach with existing motion databases for gestural interaction and performed a user evaluation with naive subjects to show its robustness to arbitrarily defined gestures. The results showed that our classification scheme has high classification accuracy for user-independent scenarios even with users who have different handedness. We believe that sparse representation of human motion will pave the way for a new generation of 3D gesture recognition systems in order to fully open the potential of natural user interfaces.



Figure 2. A participant interacting with the proposed gesture recognition system.

This work was done in collaboration with PANAMA team.

#### 7.1.2. Automatic tools for the evaluation of VR systems

**AGENT: Automatic Generation of Experimental Protocol Runtime** 

Participants: Gwendal Le Moulec, Ferran Argelaguet, Valérie Gouranton and Bruno Arnaldi

Due to the nature of Virtual Reality (VR) research, conducting experiments in order to validate the researchers' hypotheses is a must. However, the development of such experiments is a tedious and time-consuming task. We proposed in [18] to make this task easier, more intuitive and faster with a method able to describe and generate the most tedious components of VR experiments. The main objective is to let experiment designers focus on their core tasks: designing, conducting, and reporting experiments. To that end, we propose the use of Domain-Specific Languages (DSLs) to ease the description and generation of VR experiments. An analysis of published VR experiments is used to identify the main properties that characterize VR experiments. This allowed us to design AGENT (Automatic Generation of ExperimeNtal proTocols), a DSL for specifying and generating experimental protocol runtimes. AGENT allows experiment designers to design an Experimental Conditions Model (see Figure 3-left) and a Protocol Model (see Figure 3-right) in the AGENT editor. The models are then automatically compiled into code that can be integrated into VR development tools, e.g. Unity. We demonstrated the feasibility of our approach by using AGENT within two experiments.



Figure 3. Examples of "Experimental Conditions" Model (left) and "Protocol" Model (right) which are both editable with the AGENT editor.

This work was done in collaboration with DIVERSE team.

#### 7.1.3. Customer behavior and analyses in VR

## The use of immersive Virtual Reality to investigate consumer perceptions and purchase behavior toward non-standard fruits and vegetables

Participants: Jean-Marie Normand and Guillaume Moreau

With the growth of organic Fruits and Vegetables (FaVs) markets, there is now a trend in marketing research toward studies of non-standardized fruits and vegetables in stores. Yet, because of the decaying nature of FaVs, it is difficult to conduct such studies. A solution is to conduct them within a Virtual Environment (VE) (with virtual FaVs). Therefore, it is of interest to develop an approach to generate a large variety and variability

of FaVs, so one can later use them in a VE. First, we introduced a pipeline to generate a large variability of FaVs, focusing both on their shape and on their external appearance [29]. Regarding the shape, we use a semiautomated approach. A parametric Skeletal Structure and Generalized Cylinders (GCs) generates their overall shape and metaball-based techniques give them an organic aspect. Regarding their external appearance, we use a particle system approach to simulate their modifications over time. This particle system-based approach decomposes FaVs appearance changes into distinct visual characteristics producing different texture maps that can be combined.



Figure 4. Our semi-automated process using a skeletal structure and cross-sections to generate different Fruits and Vegetables (Top). An example of semi-automatically generated FaVs with different levels of deformity (Bottom).

Second, we conducted an immersive virtual reality user study aimed at investigating how customers perceive and if they would purchase non standard (i.e. misshaped) fruits and vegetables (FaVs) in supermarkets and hypermarkets [23]. Indeed, food waste is a major issue for the retail sector and a recent trend is to reduce it by selling non-standard goods. An important question for retailers relates to the FaVs' " level of abnormality " that consumers would agree to buy. However, this question cannot be tackled using " classical " marketing techniques that perform user studies within real shops since fresh produce such as FaVs tend to rot rapidly preventing studies to be repeatable or to be run for a long time. In order to overcome those limitations, we created a virtual grocery store with a fresh FaVs section where 142 participants were immersed using an Oculus Rift DK2 HMD. Participants were presented either "normal", "slightly misshaped", "misshaped" or "severely misshaped" FaVs. Results show that participants tend to purchase a similar number of FaVs whatever their deformity. Nevertheless participants' perceptions of the quality of the FaV depend on the level of abnormality.

This work was done in collaboration with Audencia Business School, Nantes, France.

#### 7.2. Physically-Based Simulation and Haptic Feedback

7.2.1. Physically-based simulation



Figure 5. Our virtual supermarket, a participant and a close-up view of the participant on the Fruits and Vegetables booth.

## Elasticity-based Clustering for Haptic Interaction with Multi-Resolution Heterogeneous Deformable Objects

Participants: Benoît Le Gouis, Maud Marchal, Bruno Arnaldi and Anatole Lécuyer

Physically-based simulation of heterogeneous objects remains a strong computational challenge for many VR applications, especially when involving haptic interaction. In [17], we introduced a novel physically-based multi-resolution approach for haptic interaction with heterogeneous deformable objects. Our method called "Elasticity-based Clustering" is based on the clustering and aggregation of elasticity inside an object, so to create large homogeneous volumes preserving important features of the initial repartition. Such a creation of large and homogeneous volumes simplifies the attribution of elasticity to the elements of the coarser geometry. We could successfully implement and test our approach within a complete and real-time haptic interaction pipeline compatible with consumer-grade haptic devices. We evaluated the performance of our approach on a large set of elasticity configurations using a perception-based quality criterion. Our results show that for 90% of studied cases our method can achieve a 6 times speedup in the simulation time with no theoretical perceptual difference.

## Real-time Target Tracking of Soft Tissues in 3D Ultrasound Images Based on Robust Visual Information and Mechanical Simulation

#### Participants: Maud Marchal

In [], we presented a real-time approach that allows tracking deformable structures in 3D ultrasound sequences. Our method consists in obtaining the target displacements by combining robust dense motion estimation and mechanical model simulation. We performed evaluation of our method through simulated data, phantom data, and real-data. Results demonstrated that this novel approach has the advantage of providing correct motion estimation regarding different ultrasound shortcomings including speckle noise, large shadows and ultrasound gain variation. Furthermore, we could show the good performance of our method with respect to state-of-theart techniques by testing on the 3D databases provided by MICCAI CLUST'14 and CLUST'15 challenges.

This work was done in collaboration with Lagadic team and IRT B-Com.



Figure 6. Overview of our elasticity-based clustering approach: (left) Based on an heterogeneous object composed of elements with different elasticity values, (center) we propose to build elasticity clusters to improve the computation time performances, (right) thus allowing haptic interaction while keeping similar perceptual sensations.

#### 7.2.2. Haptic feedback

#### Haptic Rendering of FEM-based Tearing Simulation using Clusterized Collision Detection

Participants: Benoît Le Gouis, François Lehericey, Maud Marchal, Valérie Gouranton, Bruno Arnaldi and Anatole Lécuyer

Haptic rendering of deformable phenomena remains computationally-demanding, especially when topology modifications are simulated. Within this context, the haptic rendering of tearing phenomena is under-explored as of today. In [16] we proposed a fully-functional interaction pipeline for physically-based simulation of deformable surface tearing allowing to reach haptic interactive rates. It relies on a high efficiency collision detection algorithm for deformable surface meshes, combined with an efficient FEM-based simulation of deformable surfaces enabling tearing process. We especially introduced a novel formulation based on clusters for the collision detection to improve computation time performances. Our approach was illustrated through interactive use-cases of tearing phenomena with haptic feedback, showing its ability to handle realistic rendering of deformable surface tearing on consumer-grade haptic devices.

#### FlexiFingers: Multi-Finger Interaction in VR Combining Passive Haptics and Pseudo-Haptics

Participants: Maud Marchal, Benoît Le Gouis, Ferran Argelaguet and Anatole Lécuyer

3D interaction in virtual reality often requires to manipulate and feel virtual objects with our fingers. Although existing haptic interfaces can be used for this purpose (e.g. force-feedback exoskeleton gloves), they are still bulky and expensive. We introduced a novel multi-finger device called "FlexiFingers" that constrains each digit individually and produces elastic forcefeedback [9]. FlexiFingers leverages passive haptics in order to offer a lightweight, modular, and affordable alternative to active devices. Moreover, we combined Flexifingers with a pseudo-haptic approach that simulates different levels of stiffness when interacting with virtual objects. We illustrated how this combination of passive haptics and pseudo-haptics could benefit multi-finger interaction through several use cases related to music learning and medical training. These examples suggest that our approach could find applications in various domains that require an accessible and portable way of providing haptic feedback to the fingers.

#### 7.3. Augmented Reality

#### 7.3.1. Perception in augmented reality

#### AR Feels "Softer" than VR: Haptic Perception of Stiffness in Augmented versus Virtual Reality

Participants: Yoren Gaffary, Benoît Le Gouis, Maud Marchal, Ferran Argelaguet, Anatole Lécuyer and Bruno Arnaldi



Figure 7. Our FEM-based method allows for the bimanual haptic tearing of deformable surfaces.



Figure 8. The FlexiFingers is a multi-finger device, combined with a pseudo-haptic approach, that can be used in music learning applications for instance.

Does it feel the same when you touch an object in Augmented Reality (AR) or in Virtual Reality (VR)? In [3] we studied and compared the haptic perception of stiffness of a virtual object in two situations: (1) a purely virtual environment versus (2) a real and augmented environment. We have designed an experimental setup based on a Microsoft HoloLens and a haptic force-feedback device, enabling to press a virtual piston, and compare its stiffness successively in either Augmented Reality (the virtual piston is surrounded by several real objects all located inside a cardboard box) or in Virtual Reality (the same virtual piston is displayed in a fully virtual scene composed of the same other objects). We have conducted a psychophysical experiment with 12 participants. Our results show a surprising bias in perception between the two conditions. The virtual piston is on average perceived stiffer in the VR condition compared to the AR condition. For instance, when the piston had the same stiffness in AR and VR, participants would select the VR piston as the stiffer one in 60% of cases. This suggests a psychological effect as if objects in AR would feel "softer" than in pure VR. Taken together, our results open new perspectives on perception in AR versus VR, and pave the way to future studies aiming at characterizing potential perceptual biases.



Figure 9. In our experiment, participants could interact with a virtual piston superimposed inside a real cardboard box in AR (left), and with the same piston inside a virtual replica of the box in VR (right).

#### 7.3.2. Interaction in augmented reality

## Evaluation of Facial Expressions as an Interaction Mechanism and their Impact on Affect, Workload and Usability in an AR game

#### Participants: Jean-Marie Normand and Guillaume Moreau

With the recent development of Head Mounted Display (HMD) for Virtual Reality (VR) allowing to track and recognize user's Facial Expression (FE)s in real-time, we investigated the impact that the use of FEs as an action-trigger input mechanism (e.g. a FE mapped to a single action) has on our emotional state; as well as their workload and usability compared to the use of a controller button. In [22] we developed an Augmented Reality (AR)-based memory card where the users select virtual cards using a wand and flip them using either a FE (smiling; frowning) or a Xbox controller button. The users were separated into three groups: (1) flipping the card with a smile (n = 10); (2) flipping the card with a frown (n = 8) and (3) flipping the cards with the Xbox controller button (n = 11). We did not see any significant differences between our groups in: (i) the positive and negative affect of the participants and (ii) the reported workload and usability, thus highlighting that the FEs could be used inside a HMD in the same way as a controller button.

This work was done in collaboration with the Interactive Media Lab of Keio University (Japan).

#### A Sate-of-the-Art on the Combination of Brain-Computer Interfaces and Augmented Reality



Figure 10. A participant wearing the Expression-Wear and playing our AR memory card game. Top right: The ExpressionWear embedded in the Oculus Rift. Middle and Bottom right: two states of our AR memory card game.

#### Participants: Hakim Si-Mohammed, Ferran Argelaguet and Anatole Lécuyer

We have reviewed the state-of-the art of using Brain-Computer Interfaces in combination with Augmented Reality (AR) [21]. In this work, first we introduced the field of AR and its main concepts. Second, we described the various systems designed so far combining AR and BCI categorized by their application field: medicine, robotics, home automation and brain activity visualization. Finally, we summarized and discussed the results of our survey, showing that most of the previous works made use of P300 or SSVEP paradigms with EEG in Video See-Through systems, and that robotics is a main field of application with the highest number of existing systems.

This work was done in collaboration with MJOLNIR team.

#### 7.3.3. Tracking

#### Increasing Optical Tracking Workspace of VR Applications using Controlled Cameras

Participants: Guillaume Cortes, Anatole Lécuyer

We have proposed an approach to greatly increase the tracking workspace of VR applications without adding new sensors [14]. Our approach relies on controlled cameras able to follow the tracked markers all around the VR workspace providing 6DoF tracking data. We designed the proof-of-concept of such approach based on two consumer-grade cameras and a pan-tilt head. The resulting tracking workspace could be greatly increased depending on the actuators' range of motion. The accuracy error and jitter were found to be rather limited during camera motion (resp. 0.3cm and 0.02cm). Therefore, whenever the final VR application does not require a perfect tracking accuracy over the entire workspace, we recommend using our approach in order to enlarge the tracking workspace.

This work was done in collaboration with LAGADIC team.

#### An Optical Tracking System based on Hybrid Stereo/Single-View Registration and Controlled Cameras



Figure 11. Our optical tracking prototype based on controlled cameras illustrated on a wall-sized VR application.

#### Participants: Guillaume Cortes, Anatole Lécuyer

Optical tracking is also widely used in robotics applications such as unmanned aerial vehicle (UAV) localization. Unfortunately, such systems require many cameras and are, consequently, expensive. We proposed an approach to increase again the optical tracking volume without adding cameras [13]. First, when the target becomes no longer visible by at least two cameras we propose a single-view tracking mode which requires only one camera. Furthermore, we propose to rely again on controlled cameras able to track the UAV all around the volume to provide 6DoF tracking data through multi-view registration. This is achieved by using a visual servoing scheme. The two methods can be combined in order to maximize the tracking volume. We propose a proof-of-concept of such an optical tracking system based on two consumer-grade cameras and a pan-tilt actuator and we used this approach on UAV localization.

This work was done in collaboration with LAGADIC team.

#### 7.4. Brain-Computer Interfaces

#### 7.4.1. BCI methods and techniques

#### **Designing Guiding Systems for BCI**

#### Participants: Nataliya Kosmyna and Anatole Lécuyer

The Brain–Computer Interface (BCI) community has focused the majority of its research efforts on signal processing and machine learning, mostly neglecting the human in the loop. Guiding users on how to use a BCI is crucial in order to teach them to produce stable brain patterns. In [5] we explored the instructions and feedback for BCIs in order to provide a systematic taxonomy to describe the BCI guiding systems. The purpose of our work was to give necessary clues to the researchers and designers in Human–Computer Interaction (HCI) in making the fusion between BCIs and HCI more fruitful but also to better understand the possibilities BCIs can provide to them.

#### **Towards Understanding Inverse Models in BCI**

Participants: Jussi Lindgren



Figure 12. Our optical tracking prototype based on controlled cameras and hybrid stereo/single-view registration. The tracking is used for UAV indoor localization purposes.

In the scope of the LABEX CominLabs project "SABRE", we have investigated the applicability of physiology-based source reconstruction for Brain-Computer Interfaces. The BCI interfaces leave a lot to be desired in terms of their accuracy and speed. Can source reconstruction help? We explained how the source reconstruction techniques relate to the currently mainstream machine learning methods that may recover the sources implicitly [6]. We explained the different approaches in a common linear dictionary framework and review the different ways to obtain the dictionary parameters. Our analysis suggests physiological source reconstruction may improve BCI accuracy if machine learning is not used or where it produces less optimal parameters. We considered the effect of source reconstruction on some major difficulties in BCI classification, namely information loss, feature selection and nonstationarity of the EEG. The provided analysis and discussion should help in understanding, applying, comparing and improving such techniques in the future.

#### **Cognitive Demand of BCI**

#### Participants: Andeol Evain, Ferran Argelaguet and Anatole Lécuyer

BCIs are presumably supposed to require the full attention of their users and to lose accuracy if they pay attention to another task. This assertion has been verified with several BCI paradigms (e.g. P300). But the cognitive demand of the promising SSVEP paradigm had never been specifically assessed yet. In [15] we measured the accuracy of an SSVEP-based BCI used by 26 participants in various conditions of mental workload. Our analysis revealed that surprisingly, for this type of BCI, little attention is actually needed from participants to reach optimal accuracy: participants were able to successfully perform a complex secondary task (N-back) without degrading the BCI accuracy. The same observation was made whether visual or auditive attention was solicited. These results indicate that SSVEP is a low-demanding paradigm in terms of cognitive resources, and are encouraging for its use in complex interaction settings.

This work was done in collaboration with MJOLNIR team.

#### 7.4.2. Neurofeedback

#### How to Build a Hybrid Neurofeedback Platform Combining EEG and fMRI

Participants: Marsel Mano, Lorraine Perronnet and Anatole Lécuyer

Multimodal neurofeedback estimates brain activity using information acquired with more than one neurosignal measurement technology. We have studied and described how to set up and use a hybrid platform based on simultaneous electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), then we



Figure 13. N-back task used in our study to control the level of difficuly and cognitive workload.

illustrated how to use it for conducting bimodal neurofeedback experiments in [20]. This work is intended for those willing to build a multimodal neurofeedback system, to guide them through the different steps of the design, setup, and experimental applications, and help them choose a suitable hardware and software configuration. Furthermore, we reported practical information from bimodal neurofeedback experiments conducted in our lab (see Figure 14). The platform that we presented has a modular parallel processing architecture that promotes real-time signal processing performance and simple future addition and/or replacement of processing modules. Various unimodal and bimodal neurofeedback experiments conducted in our lab showed high performance and accuracy. Currently, the platform is able to provide neurofeedback based on electroencephalography and functional magnetic resonance imaging, but the architecture and the working principles described here are valid for any other combination of two or more real-time brain activity measurement technologies.



Figure 14. Experimental platform: (A) EEG subsystem installation outside the MR room, (B) installation of the MR coil.

This work was done in collaboration with VISAGES team.

#### Unimodal Versus Bimodal EEG-fMRI Neurofeedback of a Motor Imagery Task

Participants: Lorraine Perronnet and Anatole Lécuyer

Neurofeedback is a promising tool for brain rehabilitation and peak performance training. Neurofeedback approaches usually rely on a single brain imaging modality such as EEG or fMRI. Combining these modalities for neurofeedback training could allow to provide richer information to the subject and could thus enable him/her to achieve faster and more specific self-regulation. Yet unimodal and multimodal neurofeedback have never been compared before. In [8] we introduced a simultaneous EEG-fMRI experimental protocol in which participants performed a motor-imagery task in unimodal and bimodal NF conditions (see Figure 15). With this protocol we were able to compare for the first time the effects of unimodal EEG-neurofeedback and fMRI-neurofeedback versus bimodal EEG-fMRI-neurofeedback by looking both at EEG and fMRI activations. We also proposed a new feedback metaphor for bimodal EEG-fMRI-neurofeedback that integrates both EEG and fMRI signal in a single bi-dimensional feedback (a ball moving in 2D). Such a feedback is intended to relieve the cognitive load of the subject by presenting the bimodal neurofeedback task as a single regulation task instead of two. Additionally, this integrated feedback metaphor gives flexibility on defining a bimodal neurofeedback target. Participants were able to regulate activity in their motor regions in all NF conditions. Moreover, motor activations as revealed by offline fMRI analysis were stronger during EEG-fMRIneurofeedback than during EEG-neurofeedback. This result suggests that EEG-fMRI-neurofeedback could be more specific or more engaging than EEG-neurofeedback. Our results also suggest that during EEG-fMRIneurofeedback, participants tended to regulate more the modality that was harder to control. Taken together our results shed first light on the specific mechanisms of bimodal EEG-fMRI-neurofeedback and on its addedvalue as compared to unimodal EEG-neurofeedback and fMRI-neurofeedback.



Figure 15. Real-time multimodal EEG/fMRI Neurofeedback experiment. The participant is lying in the MR tube with a 64-channel MR-compatible EEG cap. EEG and fMRI are simultaneously acquired then pre-processed with custom Matlab code. The EEG and fMRI laterality features are computed and eventually translated as a displacement of the ball on the stimulation screen, the image of which is projected on the mirror mounted on the head coil.

This work was done in collaboration with VISAGES team.

Investigating Neurophysiological Correlates of Covert Attention in Soccer Goalkeepers

Participants: Camille Jeunet, Ferran Argelaguet and Anatole Lécuyer

Soccer goalkeepers must process information from their peripheral vision at the same time they look towards the ball. This ability, committing attention to a position other than the fixation point, is called Covert Visuo-Spatial Attention or CVSA. CVSA being essential to reach high performances, it is primordial to find innovative and efficient ways of improving it. Neurofeedback, which consists in training specific brain features in order to enhance a cognitive ability, has been proven to increase attentional abilities. Also, different studies have suggested the existence of a neurophysiological marker specific to covert attention: a lateralised modulation of the alpha waves in the visual cortex. Moreover, it has been shown possible to compute this marker online, thus opening the door to a potential neurofeedback training procedure. In this view, we have proposed in a first instance to further investigate the relevance of this marker for soccer goalkeepers. The objective was here to answer the following questions: Is this marker transferrable to goalkeepers? How stable is it across athletes? Does it depend on their expertise?

This work was presented at the World Conference on Science and Soccer (Rennes, France, May 2017). It was done in collaboration with M2S Laboratory and EPFL.

#### 7.5. Cultural Heritage

#### 7.5.1. VR and AR tools for cultural heritage

#### **EvoluSon: Walking Through an Interactive History of Music**

Participants: Ronan Gaugne, Florian Nouviale and Valérie Gouranton

The EvoluSon project [4] proposes an immersive experience where the spectator explores an interactive visual and musical representation of the main periods of the history of Western music. The musical content is constituted of original musical compositions based on the theme of Bach's Art of Fugue to illustrate eight main musical eras from Antiquity to the contemporary epoch. The EvoluSon project contributes at the same time to the usage of VR for intangible culture representation and to interactive digital art that puts the user at the centre of the experience. The EvoluSon project focuses on music through a presentation of the history of Western music, and uses virtual reality to valorise the different pieces through the ages. The user is immersed in a coherent visual and sound environment and can interact with both modalities (see Figure 16).

This project was done in collaboration with the Research Laboratory on Art and Music of University Rennes 2.

#### **Immersive Point Cloud Manipulation for Cultural Heritage Documentation**

#### Participants: Jean-Baptiste Barreau, Ronan Gaugne and Valérie Gouranton

Virtual reality combined with 3D digitisation allows to immerse archaeologists in 1:1 copies of monuments and sites. However, scientific communication of archaeologists is based on 2D representations of the monuments they study. In [2] we proposed a virtual reality environment with an innovative cutting-plan tool to dynamically produce 2D cuts of digitized monuments. A point cloud is the basic raw data obtained when digitizing cultural heritage sites or monuments with laser scans or photogrammetry. These data represent a rich and faithful record provided that they have adequate tools to exploit them. Their current analyses and visualizations on PC require software skills and can create ambiguities regarding the architectural dimensions. We conceived a toolbox to explore and manipulate such data in an immersive environment, and to dynamically generate 2D cutting planes usable for cultural heritage documentation and reporting (see Figure 17).

#### MAAP Annotate: When Archaeology meets Augmented Reality for Annotation of Megalithic Art

#### Participants: Jean-Marie Normand

Megalithic art is a spectacular form of symbolic representation found on prehistoric monuments. Carved by Europe's first farmers, this art allows an insight into the creativity and vision of prehistoric communities. As examples of this art continue to fade, it is increasingly important to document and study these symbols. In [11] we introduced MAAP Annotate, a Mixed Reality annotation tool from the Megalithic Art Analysis Project (MAAP). It provides an innovative method of interacting with megalithic art, combining cross-disciplinary research in digital heritage, 3D scanning and imaging, and augmented reality. The development of the tool



Figure 16. EvoluSon application in Immersia, Middle-Age era.



Figure 17. Immersive manipulation of the point cloud of the "Salle du Jeu de Paume" of Rennes, in Immersia room.

is described, alongside the results of an evaluation carried out on a group of archaeologists from University College Dublin, Ireland. It is hoped that such tools will enable archaeologists to collaborate worldwide, and non-specialists to experience and learn about megalithic art (see Figure 18).

This work was done in collaboration with the School of Computer Science and Informatics and the School of Archaeology of University College Dublin (UCD).



Figure 18. A real Irish megalith engraved with petroglyphs (Top-left) and a participant using the HoloLens to annotate the 3D scan of the megalith (Top-right). Overview of the MAAP Annotate user interface (Middle row), and two examples of manual AR annotations (Bottom row).

#### 7.5.2. Multi-modal images and 3D printing for cultural heritage

#### Physical Digital Access Inside Archaeological Material

Participants: Théophane Nicolas, Ronan Gaugne and Valérie Gouranton

Traditionally, accessing the interior of an artefact or an archaeological material is a destructive activity. We proposed an alternative non-destructive technique, based on a combination of medical imaging and advanced transparent 3D printing. Our approach proposes combining a computed tomography (CT) scan and advanced 3D printing to generate a physical representation of an archaeological artefact or material [7]. This project is conducted with archaeologists from Inrap and computer scientists from Inria-IRISA. The goal of the project is to propose innovative practices, methods and tools for archaeology based on 3D digital techniques. We notably proposed a workflow (see Figure 19) where the CT scan images are used to produce volume and surface 3D data which serve as a basis for new evidence usable by archaeologists. This new evidence can be observed in interactive 3D digital environments or through physical copies of internal elements of the original material.



This work was done in collaboration with Inrap.

Figure 19. Our workflow for combining a computed tomography (CT) scan and advanced 3D printing to generate a physical representation of an archaeological artefact.

#### Combining CT-scan, Photogrammetry, 3D Printing and Mixed Reality

Participants: Théophane Nicolas, Ronan Gaugne, Bruno Arnaldi and Valérie Gouranton

Archaeological artefacts and the sediments that contain them constitute the sometimes tenuous evidence that requires analysis, preservation and showcasing. Different methods of digital analysis that provide non destructive solutions to preserve, analyse and showcase archaeological heritage have been developed over recent years. However these techniques are often restricted to the visible surface of the objects, monuments or sites. The techniques used in medical imaging are more and more frequently used in archaeology as they give non destructive access to the artefacts' internal and often fragile structure. This use is mostly limited to a simple visualisation. The information obtained by CT-scan is transcribed in a visual manner and its inherent detail can be used much more widely in the domain of the latest 3D technologies such as virtual reality, augmented reality, multimodal interactions and additive manufacturing. In combining these medical imaging techniques, it becomes possible to identify and scientifically analyse by efficient and non destructive methods non visible objects, to assess their fragility and their state of preservation. It is also possible to assess the restoration of a corroded artefact, to visualise, to analyse and to physically manipulate an inaccessible or fragile object (CT, 3D printing) and to observe the context of our hidden archaeological heritage (virtual reality, augmented reality or mixed, 3D). The development of digital technologies will hopefully lead to a democratisation of this type of analysis. We could illustrate our approach using the study of several artefacts from the recent excavation of the

Warcq chariot burial (Ardennes, France). We notably presented in [28] a physical interaction with inaccessible objects based on a transparent 3D printing of a horse's cranium (see Figure 20).

This work was done in collaboration with Inrap, Image ET and University Paris 1.



Figure 20. Original horse cranium (Top-left) and CT-scan of the cranium (Top-right), volume rendering from the CT scan (Bottom-left) and then transparent 3D printing of the same object (Bottom-right).

#### A Multimodal Digital Analysis of a Mesolithic Clavicle: Preserving and Studying the Oldest Human Bone in Brittany

#### Participants: Jean-Baptiste Barreau, Ronan Gaugne and Valérie Gouranton

The oldest human bone of Brittany was dug up from the mesolithic shell midden of Beg-er-Vil in Quiberon and dated about 8200 years. The low acid soils of these dump area represent exceptional sedimentary conditions. For these reasons, but also because these bones have a very particular patrimonial and symbolic value, their study goes altogether with concerns of conservation and museographic presentation. The clavicle is constituted of two pieces discovered separately at a one meter distance from each other. The two pieces match, so it can be assemble in a single fragment of approximately 7 centimeters. Cut-marks are clearly visible on the surface of these bones. They are bound to post-mortem processing which it is necessary to better qualify. The clavicle was studied through a process that combines advanced 3D image acquisition, 3D processing, and 3D printing with the goal to provide relevant support for the experts involved [24]. The bones were first scanned with a CT scan, and digitized with photogrammetry in order to get a high quality textured model. The CT scan appeared to be insufficient for a detailed analysis. The study was thus completed with a  $\mu$ CT providing a very accurate 3D model of the bone. Several 3D-printed copies of the collarbone were produced in order to constitute tangible support easy to annotate for sharing between the experts involved in the study. The 3D models generated from  $\mu$ CT and photogrammetry, were combined to provide an accurate and detailed 3D model. This model was used to study desquamation and the different cut marks. These cut marks were also studied with traditional binoculars and digital microscopy. This last technique allowed characterizing the cut marks, revealing a probable meat cutting process with a flint tool (see Figure 21). This work of crossed analyses allowed to document a fundamental patrimonial piece, and to ensure its preservation.

This work was done in collaboration with Inrap, UMR CREAAH, CNRS-INE and Université Paris 1.



*Figure 21. Digital microscopy of the clavicle (Top-left), and then detail of the cut marks in the digital model (Top-right), and annotated 3D printed clavicle (Bottom).* 

#### 7.5.3. Generating 3D data for cultural heritage

#### 3D Reconstruction of the Fortified Entrance of the Citadel of Aleppo from a few Sightseeing Photos

#### Participants: Jean-Baptiste Barreau and Ronan Gaugne

Built at the beginning of the 16th century by the final Mamluk sultan Al-Achraf Qânsûh Al-Ghûrî, the entrance to the Citadel of Aleppo was particularly affected by an earthquake in 1822, bombings during the Battle of Aleppo in August 2012, and a collapse of ramparts due to an explosion in July 2015. Even if compared to other Syrian sites, there are still enough vestiges to grasp the initial architecture, the civil war situation makes extremely difficult any "classic" process of digitization by photogrammetry or laser scanning. On this basis, we proposed in [25] a process to produce a 3D model "as relevant as possible" only from a few sightseeing photographs. This process combines fast 3D sketching by photogrammetry, 3D modeling and texture mapping and relies on a corpus based on pictures available on the net. Furthermore, it has the advantage to be applicable to destroyed monuments if sufficient pictures are available (see Figure 22).

This work was done in collaboration with UMR CReAAH and Inrap.

#### Raising the Elevations of a Megalithic Architecture: Methodological Developments

Participants: Jean-Baptiste Barreau, Quentin Petit and Ronan Gaugne



Figure 22. Rendering of the 3D model of Aleppo entrance.

Elevations have been little studied during explorations of megalithic architectures. For the past ten years, interest in these elevations has been growing in western France, particularly with the application of archeology of buildings and its tools to study them. Megalithic architecture, however, has its own characteristics that make manual surveys difficult. The first step presented in [26] was to acquire a 3D model, precise and manageable, of the whole architecture. Photogrammetry was tested, however the small space made it difficult to photograph. Laser scanner scanning has therefore been preferred. From the cloud of points obtained, a computer processing protocol was developed in order to obtain 2D images of the elevations according to the desired views. On these, a stone-to-stone design is possible in the laboratory and rectifiable directly in the field thanks to the use of a tablet computer. Our method has therefore met accessibility constraints. Above all, it allowed to increase the time devoted to the observation of ground, with a final result identical to a manual survey.

This work was done in collaboration with UMR CReAAH and SED Rennes.

## 8. Bilateral Contracts and Grants with Industry

#### 8.1. Bilateral Contracts with Industry

#### 8.1.1. Mensia Technologies

Participants: Anatole Lécuyer, Jussi Tapio Lindgren.

Mensia Technologies is an Inria start-up company created in November 2012 as a spin-off of Hybrid team. Mensia is focused on wellness and healthcare applications emerging from the BCI and Neurofeedback technologies. The Mensia startup should benefit from the team's expertise and of valuable and proprietary BCI research results. Mensia is based in Rennes and Paris. Anatole Lécuyer and Yann Renard (former Inria expert engineer who designed the OpenViBE software architecture and was involved in team projects for 5 years) are co-founders of Mensia Technologies together with CEO Jean-Yves Quentel.

The on-going contract between Hybrid and Mensia started in November 2013 and supports the transfer of several softwares designed by Hybrid team ("OpenViBE", "StateFinder") related to our BCI activity to Mensia Technologies for multimedia or medical applications of Mensia.



Figure 23. Inside the cairn of Barnenez in Immersia room.

#### 8.2. Bilateral Grants with Industry

#### 8.2.1. Technicolor

Participants: Antoine Costes, Anatole Lécuyer, Ferran Argelaguet.

This grant started in December 2015. It supports Antoine Costes's CIFRE PhD program with Technicolor company on "Haptic Texturing".

#### 8.2.2. Realyz

Participants: Guillaume Cortes, Anatole Lécuyer.

This grant started in December 2015. It supports Guillaume Cortes's CIFRE PhD program with Realyz company on "Improving tracking in VR".

#### 8.2.3. VINCI Construction

Participants: Anne-Solène Dris-Kerdreux, Bruno Arnaldi, Valérie Gouranton.

This grant started in November 2015. It supports Anne-Solene Dris-Kerdreux's CIFRE PhD program with Vinci company on "Training in VR for construction applications".

#### 8.2.4. Orange Labs

Participants: Guillaume Bataille, Bruno Arnaldi, Valérie Gouranton.

This grant started in October 2017. It supports Guillaume Bataille's PhD program with Orange Labs company on "Natural Interactions with IoT using VR/AR".

## 9. Partnerships and Cooperations

#### 9.1. Regional Initiatives

#### 9.1.1. Labex Cominlabs SUNSET

Participants: Bruno Arnaldi, Guillaume Claude, Gautier Picard, Valérie Gouranton [contact].

SUNSET is a 4-year Labex Cominlabs project (2016-2020). SUNSET partners are MediCIS-LTSI (coordinator), Hybrid, Hycomes (IRISA/Inria), and CHU Rennes. SUNSET aims at developing an innovative training software suite based on immersive and collaborative virtual reality technology for training and evaluating non-technical skills. This approach will be implemented and evaluated in the context of training neurosurgical scrub nurses. We will notably integrate methods and systems developed in the S3PM project (see bellow). By relying on Human Factors approaches, the project also addresses training and evaluation of interpersonal skills. Whereas the developed technologies and approaches will be generic and adaptable to any surgical specialty, the project will evaluate the developed system within training sessions performed with scrub nurses. We ambition to propose novel approaches for surgical non-technical skill learning and assessment, and to install the developed training factory at the University Hospital of Rennes, and evaluate it with real-scale user studies.

#### 9.1.2. Labex Cominlabs S3PM

Participants: Bruno Arnaldi, Guillaume Claude, Valérie Gouranton [contact].

S3PM ("Synthesis and Simulation of Surgical Process Models") is a 4-year Labex Cominlabs project (2013-2017). S3PM partners are MediCIS-LTSI (coordinator), Hybrid, Hycomes (IRISA/Inria), and CHU Rennes. The objective of S3PM is to propose a solution for the computation of surgical procedural knowledge models from recordings of individual procedures, and their execution. The goal of the Hybrid team is to propose and use new models for collaborative and interactive virtual environments for procedural training. The Hybrid team also works on the creation of a surgical training application in virtual reality, exposing the different contributions.

#### 9.1.3. Labex Cominlabs HEMISFER

Participants: Anatole Lécuyer [contact], Marsel Mano, Lorraine Perronnet.

**HEMISFER** is a 4-year project (2013-2017) funded by Labex CominLabs. It involves 4 Inria/IRISA teams (Hybrid, Visages (lead), Panama, Athena) and 2 medical centers: the Rennes Psychiatric Hospital (CHGR) and the Reeducation Department of Rennes Hospital (CHU Pontchaillou). The goal of HEMISFER is to make full use of neurofeedback paradigm in the context of rehabilitation and psychiatric disorders. The major breakthrough will come from the use of a coupling model associating functional and metabolic information from Magnetic Resonance Imaging (fMRI) to Electro-encephalography (EEG) to "enhance" the neurofeedback protocol. Clinical applications concern motor, neurological and psychiatric disorders (stroke, attention-deficit disorder, treatment-resistant mood disorders, etc).

#### 9.1.4. Labex Cominlabs SABRE

Participants: Anatole Lécuyer [contact], Jussi Tapio Lindgren, Nataliya Kos' Myna.

SABRE is a 3-year project (2014-2017) funded by Labex CominLabs. It involves 1 Inria/IRISA team (Hybrid) and 2 groups from TELECOM BREST engineering school. The goal of SABRE is to improve computational functionnalities and power of current real-time EEG processing pipelines. The project will investigate innovative EEG solution methods empowered and speeded-up by ad-hoc, transistor-level, implementations of their key algorithmic operations. A completely new family of fully-hardware-integrated, new computational EEG imaging methods will be developed that are expected to speed up the imaging process of an EEG device of several orders of magnitude in real case scenarios.

#### 9.1.5. IRT b<>com

Participants: Bruno Arnaldi [contact], Valérie Gouranton, Maud Marchal.

**b**<>com is a French Institute of Research and Technology (IRT). The main goal of this IRT is to fasten the development and marketing of tools, products and services in the field of digital technologies. Our team has already collaborated with b<>com within two 3-year projects: ImData (on "Immersive Interaction") and GestChir (on "Augmented Healthcare") which both ended in 2016. A new 3-year project called NeedleWare (on "Augmented Healthcare") has started on October 2016.

#### 9.1.6. CNPAO Project

Participants: Valérie Gouranton [contact], Jean-Baptiste Barreau, Ronan Gaugne.

CNPAO ("Conservatoire Numérique du Patrimoine Archéologique de l'Ouest") is an on-going research project partially funded by the Université Européenne de Bretagne (UEB) and Université de Rennes 1. It involves IRISA/Hybrid and CReAAH. The main objectives are: (i) a sustainable and centralized archiving of 2D/3D data produced by the archaeological community, (ii) a free access to metadata, (iii) a secure access to data for the different actors involved in scientific projects, and (iv) the support and advice for these actors in the 3D data production and exploration through the latest digital technologies, modeling tools and virtual reality systems. This project involves a collaboration with Quentin Petit (SED Inria Rennes).

#### 9.1.7. Imag'In CNRS IRMA

**Participants:** Bruno Arnaldi, Jean-Baptiste Barreau, Ronan Gaugne, Valérie Gouranton [contact], Théophane Nicolas.

The IRMA project is an Imag'In project funded by CNRS which aims at developping innovative methodologies for research in the field of cultural heritage based on the combination of medical imaging technologies and interactive 3D technologies (virtual reality, augmented reality, haptics, additive manufacturing). It relies on close collaborations with the National Institute of Preventive Archaeological Research (Inrap), the Research Center Archaeology, and History Archéosciences (CReAAH UMR 6566) and the company Image ET. The developed tools are intended for cultural heritage professionals such as museums, curators, restorers, and archaeologists. We focus on a large number of archeological artefacts of different nature, and various time periods (Paleolithic, Mesolithic, and Iron Age Medieval) from all over France. We can notably mention the oldest human bones found in Brittany (clavicle Beg Er Vil), a funeral urn from Trebeurden (22), or a Bronze Cauldron from a burial of the Merovingian necropolis "Crassés Saint-Dizier" (51). This project involves a collaboration with Quentin Petit (SED Inria Rennes) and Grégor Marchand (CNRS/UMR CREAAH).

#### **9.2.** National Initiatives

#### 9.2.1. ANR-FRQSC INTROSPECT

**Participants:** Valérie Gouranton [contact], Bruno Arnaldi, Ronan Gaugne, Jean-Baptiste Barreau, Flavien Lecuyer.

**INTROSPECT** is a 3-year project funded by French ANR and "Fonds de Recherche Société et Culture" (FRQSC) from Quebec region, Canada. The collaboration involves researchers in computer science and archeology from France and Canada : Hybrid (Inria-IRISA), CReAAH, Inrap, company Image ET, University Laval and INRS-ETE. INTROSPECT aims to develop new uses and tools for archaeologists that facilitate access to knowledge through interactive numerical introspection methods that combine computed tomography with 3D visualization technologies, such as Virtual Reality, tangible interactions and 3D printing. The scientific core of the project is the systematization of the relationship between the artefact, the archaeological context, the digital object and the virtual reconstruction of the archaeological context that represents it and its tangible double resulting from the 3D printing. This axiomatization of its innovative methods makes it possible to enhance our research on our heritage and to make use of accessible digital means of dissemination. This approach changes from traditional methods and applies to specific archaeological problems. Several case studies will be studied in various archaeological contexts on both sides of the Atlantic. Quebec museums are partners in the project to spread our results among the general public.

#### 9.2.2. Ilab CertiViBE

Participants: Anatole Lécuyer [contact], Jussi Tapio Lindgren, Thierry Gaugry, Cédric Riou.

CertiViBE is a 2-year "Inria Innovation Lab" (2015-2017) funded by Inria for supporting the development of OpenViBE software, and notably its evolution in order to enable and fasten the medical transfer and the medical certification of products based on OpenViBE. This joint lab involves two partners: Hybrid and Mensia Technologies startup company. The project aims at setting up a quality environment, and developping a novel version of the software wich should comply with medical certification rules.

#### 9.2.3. IPL BCI-LIFT

**Participants:** Anatole Lécuyer [contact], Jussi Tapio Lindgren, Hakim Si Mohammed, Lorraine Perronnet, Nataliya Kos'Myna.

**BCI-LIFT** is a 4-year "Inria Project Lab" initiative (2015-2019) funded by Inria for supporting a national research effort on Brain-Computer Interfaces. This joint lab involves several Inria teams: Hybrid, Potioc, Athena, Neurosys, Mjolnir, Demar; as well as external partners: INSERM-Lyon, and INSA Rouen. This project aims at improving several aspects of Brain-Computer Interfaces: learning and adaptation of BCI systems, user interfaces and feedback, training protocols, etc.

#### 9.2.4. ATT CONSORVIBE

Participants: Anatole Lécuyer [contact], Jussi Tapio Lindgren [contact].

CONSORVIBE is a 6-month ATT Inria Project funded by Inria for supporting a prospective effort and the feasibility study of building a consortium of partners dedicated to the sustainability and promotion of the OpenViBE software.

#### 9.3. European Initiatives

#### 9.3.1. FP7 & H2020 Projects

9.3.1.1. HAPPINESS

Title: HAptic Printed Patterned INtErfaces for Sensitive Surface

Programm: H2020

Duration: January 2015 - December 2017

Coordinator: CEA (France)

Partners:

Arkema France (France)

Robert Bosch (Germany)

Commissariat A L'Energie Atomique et Aux Energies Alternatives (France)

Fundacion Gaiker (Spain)

Integrated Systems Development S.A. (Greece)

University of Glasgow (United Kingdom)

Walter Pak SL (Spain)

Inria contact: Nicolas Roussel and Anatole Lécuyer

The Automotive HMI (Human Machine Interface) will soon undergo dramatic changes, with large plastic dashboards moving from the 'push-buttons' era to the 'tactile' era. User demand for aesthetically pleasing and seamless interfaces is ever increasing, with touch sensitive interfaces now commonplace. However, these touch interfaces come at the cost of haptic feedback, which raises concerns regarding the safety of eyeless interaction during driving. The HAPPINESS project intends to address these concerns through technological solutions, introducing new capabilities for haptic feedback on these interfaces. The main goal of the HAPPINESS project is to develop a smart conformable surface able to offer different tactile sensations via the development of a Haptic Thin and Organic Large Area Electronic technology (TOLAE), integrating sensing and feedback capabilities, focusing on user requirements and ergonomic designs. To this aim, by gathering all the value chain actors (materials, technology manufacturing, OEM integrator) for application within the automotive market, the HAPPINESS project will offer a new haptic Human-Machine Interface technology, integrating touch sensing and disruptive feedback capabilities directly into an automotive dashboard. Based on the consortium skills, the HAPPINESS project will demonstrate the integration of Electro-Active Polymers (EAP) in a matrix of mechanical actuators on plastic foils. The objectives are to fabricate these actuators with large area and cost effective printing technologies and to integrate them through plastic molding injection into a small-scale dashboard prototype. We will design, implement and evaluate new approaches to Human-Computer Interaction on a fully functional prototype that combines in packaging both sensors and actuator foils, driven by custom electronics, and accessible to end-users via software libraries, allowing for the reproduction of common and accepted sensations such as Roughness, Vibration and Relief. In this project, the role of Hybrid team is to design user studies on tactile perception, and study innovative usages of the technologies developed in HAPPINESS.

#### 9.3.1.2. IMAGINE

Title: IMAGINE - Robots Understanding Their Actions by Imagining Their Effects

Programm: H2020

Duration: January 2017 - December 2020

Coordinator: Univ. Innsbruck (Austria)

Partners:

Univ. Innsbruck (Austria)

Univ. Göttingen (Germany)

Karlsruhe Institute of Technology (Germany)

**INSA Rennes (France)** 

Institute of Robotics and Industrial Informatics (Spain)

Univ. Bogazici (Turkey)

Electro Cycling (Germany)

Inria contact: Maud Marchal

Today's robots are good at executing programmed motions, but they do not understand their actions in the sense that they could automatically generalize them to novel situations or recover from failures. **IMAGINE** seeks to enable robots to understand the structure of their environment and how it is affected by its actions. "Understanding" here means the ability of the robot (a) to determine the applicability of an action along with parameters to achieve the desired effect, and (b) to discern to what extent an action succeeded, and to infer possible causes of failure and generate recovery actions. The core functional element is a generative model based on an association engine and a physics simulator. "Understanding" is given by the robot's ability to predict the effects of its actions, before and during their execution. This allows the robot to choose actions and parameters based on their simulated performance, and to monitor their progress by comparing observed to simulated behavior. This scientific objective is pursued in the context of recycling of electromechanical appliances.
Current recycling practices do not automate disassembly, which exposes humans to hazardous materials, encourages illegal disposal, and creates significant threats to environment and health, often in third countries. IMAGINE will develop a TRL-5 prototype that can autonomously disassemble prototypical classes of devices, generate and execute disassembly actions for unseen instances of similar devices, and recover from certain failures. For robotic disassembly, IMAGINE will develop a multi-functional gripper capable of multiple types of manipulation without tool changes. IMAGINE raises the ability level of robotic systems in core areas of the work programme, including adaptability, manipulation, perception, decisional autonomy, and cognitive ability. Since only one-third of EU e-waste is currently recovered, IMAGINE addresses an area of high economical and ecological impact.

#### 9.4. International Research Visitors

#### 9.4.1. Visits of International Scientists

This year, Hybrid team has welcomed for short periods:

- Gabriel Cirio, Universidad Rey Juan Carlos Madrid (Spain), May 2017
- Victoria Interrante, University of Minnesota (US), December 2017
- Geneviève Treyvaud and Pierre Francus, INRS (Canada), November 2017

#### 9.4.2. Visits to International Teams

Ronan Gaugne and Valérie Gouranton made a short stay at University Laval (Canada) in August 2017

## **10.** Dissemination

#### **10.1. Promoting Scientific Activities**

#### 10.1.1. Scientific events organisation

#### 10.1.1.1. General chair, Scientific chair

- Bruno Arnaldi was General Chair of IGRV 2017 (journées du GDR IG-RV, Journées de l'AFRV, Journées de l'AFIG), Rennes, France.
- Anatole Lécuyer and Guillaume Moreau were General co-Chairs of IEEE Symposium on Mixed and Augmented Reality 2017 (IEEE ISMAR), Nantes, France.
- Maud Marchal was Program Chair of IEEE Symposium on 3D User Interfaces 2017 (IEEE 3DUI).
- Jean-Marie Normand was Deputy General Chair of IEEE Symposium on Mixed and Augmented Reality 2017 (IEEE ISMAR), Nantes, France.

#### 10.1.1.2. Member of the organizing committees

- Anatole Lécuyer was Member of the organization committee of IGRV 2017.
- Ferran Argelaguet was Member of the organization committee of ISMAR 2017 and IGRV 2017.
- Ronan Gaugne was Member of the organization committee of ISMAR 2017 and IGRV 2017.
- Valérie Gouranton was Member of the steering committee of ISMAR 2017 and IGRV 2017.
- Maud Marchal was Member of the organization committee of IGRV 2017.
- Florian Nouviale was Member of the organization committee of ISMAR 2017 and IGRV 2017.

#### 10.1.2. Scientific events selection

10.1.2.1. Member of the conference program committees

• Anatole Lécuyer was Member of the conference program committee of IEEE Conference on Virtual Reality 2017.

- Ferran Argelaguet was Member of the conference program committee of IEEE Symposium on 3D User Interfaces 2017, and ACM Symposium on Spatial User Interfaces 2017 and IHM 2017.
- Maud Marchal was Member of the conference program committee of Eurographics 2017, the best paper committee of "Journées Françaises de l'Informatique Graphique" 2017, the best Phd award committee of GDR-IGRV/EGFR/AFRV.
- Guillaume Moreau was Member of the program committee of IEEE Symposium on 3D User Interfaces 2017, Area Chair of IAPR Machine Vision Applications 2017.
- Jean-Marie Normand was Member of the program committee of Augmented Human 2017, Area Chair of IAPR Machine Vision Applications 2017.

10.1.2.2. Reviewer

- Ferran Argelaguet was Reviewer for IEEE VR 2017, ACM CHI 2017, ACM VRST 2017, ACM UIST 2017, WorldHaptics 2017.
- Maud Marchal was Reviewer for IEEE VR 2017, Eurographics 2017, IEEE ICRA 2017.
- Valérie Gouranton was Reviewer for IEEE VR 2017.
- Guillaume Moreau was Reviewer for IEEE 3DUI 2017, IEEE VR 2017, ACM AH 2017.
- Jean-Marie Normand was Reviewer for IEEE 3DUI 2017, IEEE VR 2017, IEEE ISMAR 2017, ACM CHI 2017, ACM MM 2017.

#### 10.1.3. Journal

10.1.3.1. Member of the editorial boards

- Anatole Lécuyer is Associate Editor of the IEEE Transactions on Visualization and Computer Graphics (since Dec 2017), Frontiers in Virtual Environments, and Presence journals. He is also Guest Editor of a Special Issue of IEEE Computer Graphics and Applications on "Virtual and Augmented Reality".
- Ferran Argelaguet is Review Editor of Frontiers in Virtual Environments.
- Maud Marchal is Review Editor of Frontiers in Virtual Environments.
- Guillaume Moreau is Associate Editor of a Special Issue of the Transactions of the Institute of Electronics, Information and Communication Engineers and for Applied Sciences.
- Jean-Marie Normand is Review Editor of Frontiers in Virtual Environments.

#### 10.1.3.2. Reviewer - Reviewing activities

- Ferran Argelaguet was Reviewer for ACM Transactions on Graphics, IEEE Transactions on Visualization and Computer Graphics, International Journal on Human Computer Studies.
- Ronan Gaugne was Reviewer for Digital Applications in Archaeology and Cultural Heritage.
- Maud Marchal was Reviewer for IEEE Transactions on Visualization and Computer Graphics, IEEE Transactions on Haptics, The Visual Computer and Computers and Graphics Journal.
- Guillaume Moreau was reviewer for IEEE Transactions on Visualization and Computer Graphics and Computers, Environments and urban Systems.
- Jean-Marie Normand was Reviewer for Frontiers in Robotics and AI, IEEE Computer Graphics and Applications, IEEE Transactions on Visualization and Computer Graphics.

#### 10.1.4. Invited talks

- Anatole Lécuyer was Keynote Speaker at EuroVR 2017.
- Guillaume Moreau was Keynote Speaker at "Journées de l'Association Française d'Histotechnologie".

#### 10.1.5. Leadership within the scientific community

- Bruno Arnaldi is Member of Executive Committee of AFRV (French Association for Virtual Reality).
- Ronan Gaugne is Member of the Selection and Validation Committee for the French cluster "Pôle Images et Réseaux"
- Valérie Gouranton is Member of Executive Committee of AFRV.
- Maud Marchal is Member of Executive Committee of Eurographics French Chapter.

#### 10.1.6. Scientific expertise

- Valérie Gouranton was Member of a selection committee for French ANR.
- Ronan Gaugne was Expert for "Atlanstic2020 Région Pays de la Loire".
- Maud Marchal was Expert for "Haute Ecole Specialisée de Suisse Occidentale" (Switzerland).
- Jean-Marie Normand was Expert for French ANR.

#### 10.1.7. Research administration

- Bruno Arnaldi is Deputy Director of IRISA, and co-Head of the Scientific Council of University of Rennes (ENS Rennes, ENSC Rennes, IEP Rennes, INSA Rennes, University Rennes 1, and University Rennes 2).
- Maud Marchal is Co-Head of the Master of "Research in Computer Science" (SIF) at University Rennes 1.
- Valérie Gouranton is Head of cross-cutting Axis "Art, Heritage & Culture" at IRISA.
- Jean-Marie Normand is Head of the Minor "Virtual Reality" at ECN.

#### **10.2. Teaching - Supervision - Juries**

Anatole Lécuyer:

Master MNRV: "Haptic Interaction", 9h, M2, ENSAM, Laval, FR

Ecole Centrale de Nantes : "Haptic Interaction and Brain-Computer Interfaces", 4.5h, M1-M2, Ecole Centrale de Nantes, FR

Master SIBM: "Haptic and Brain-Computer Interfaces", 4.5h, M2, University of Rennes 1, FR

Master CN: "Haptic Interaction and Brain-Computer Interfaces", 9h, M1 and M2, University of Rennes 2, FR

Master VRI: "Pseudo-Haptics and Brain-Computer Interfaces", 6h, M2, University of Rennes 1, FR

#### Bruno Arnaldi:

Master INSA Rennes: "VRI: Virtual Reality and Multi-Sensory Interaction Course", 4h, M2, INSA Rennes, FR

Master INSA Rennes: "CG: Computer Graphics", 10h, M2, INSA Rennes, FR

Master INSA Rennes: "Virtual Reality", courses 6h, projects 16h, M1 and M2, INSA Rennes, FR

Master INSA Rennes: Projects on "Virtual Reality", 20h, M1, INSA Rennes, FR

#### Ferran Argelaguet:

Master STS Informatique MITIC: "Techniques d'Interaction Avancées", 26h, M2, ISTIC, University of Rennes 1, FR

Master INSA Rennes: "Modeling and Engineering for Biology and Health Applications", 12h, M2, INSA Rennes, FR

Maud Marchal:

Master of Research in Computer Science: "Haptic rendering and physically-based simulation", 4h, M2, University of Rennes 1, FR

Master INSA Rennes: "Computer Graphics", 26h, M1 and responsible of this lecture, INSA Rennes, FR

Master INSA Rennes: "Modeling and Engineering for Biology and Health Applications", 48h, M2 and responsible of this lecture, INSA Rennes, FR

Master SIBM: "Biomedical simulation", 3h, M2, University of Rennes 1, FR

#### Valérie Gouranton:

Licence: "Introduction to Virtual Reality", 22h, L2 and responsible of this lecture, INSA Rennes, FR

Licence: Project on "Virtual Reality", 16h, L3 and responsible of this lecture, INSA Rennes, FR Master INSA Rennes: "Virtual Reality", 16h, M2, INSA Rennes, FR

Master INSA Rennes: Projects on "Virtual Reality", 20h, M1, INSA Rennes, FR

Master CN: "Virtual Reality", 6h, M1, University of Rennes 2, FR

#### Ronan Gaugne:

INSA Rennes: Projects on "Virtual Reality", 47h, L3/M1/M2, Insa Rennes, FR

Master CN: "Virtual Reality", 3h, M1, University of Rennes 2, FR

Jean-Marie Normand:

Virtual Reality Minor, "Computer Graphics", 24h, M1/M2, Ecole centrale de Nantes, FR

Virtual Reality Minor, "Fundamentals of Virtual Reality", 14h, M1/M2, Ecole centrale de Nantes, FR

Virtual Reality Minor, "Computer Vision anf Augmented Reality", 24h, M1/M2, Ecole centrale de Nantes, FR

Virtual Reality Minor, "Projects on Virtual Reality", 20h, M1/M2, Ecole centrale de Nantes, FR

#### 10.2.1. Supervision

10.2.1.1. PhD (defended)

- Jean-Baptiste Barreau, "Techniques of production, exploration and analysis of virtual archaeological environments", INSA Rennes, July 10th 2017, Supervised by Valérie Gouranton and Bruno Arnaldi
- Lorraine Perronnet, "Combining electroencephalography and functional magnetic resonance imaging for neurofeedback", University of Rennes 1, January 8th, 2017, Supervised by Anatole Lécuyer, Fabien Lotte (Potioc, Inria), Maureen Clerc (Athena, Inria) and Christian Barillot (Visages, Inria)
- Benoit Le Gouis, "Contribution to the study of haptic rendering and perception of virtual deformable objects", INSA Rennes, November 21th 2017, Supervised by Bruno Arnaldi, Maud Marchal and Anatole Lécuyer

#### 10.2.1.2. PhD (in progress)

- Gwendal Le Moulec, "Automatic generation of VR applications", Started in October 2015, Supervised by Valérie Gouranton, Bruno Arnaldi and Arnaud Blouin (Diverse, Inria)
- Anne-Solène Dris-Kerdreux, "Training in virtual reality for construction applications", Started in November 2015, Supervised by Valérie Gouranton and Bruno Arnaldi
- Antoine Costes, "Haptic texturing", Started in November 2015, Supervised by Anatole Lécuyer, Philippe Guillotel (Technicolor), Fabien Danieau (Technicolor) and Ferran Argelaguet
- Guillaume Cortes, "Improving tracking in VR", Started in November 2015, Supervised by Anatole Lécuyer and Eric Marchand (Lagadic, Inria)

- Hakim Si-Mohammed, "BCI and HCI", Started in October 2016, Supervised by Anatole Lécuyer, Géry Casiez (Mjolnir, Inria), Nicolas Roussel (Mjolnir, Inria) and Ferran Argelaguet
- Gautier Picard, "Collaborative VR", Started in October 2016, Supervised by Valérie Gouranton, Bernard Gibaud (Inserm) and Bruno Arnaldi
- Hadrien Gurnel, "Prise en compte de la déformation d'organe pour l'assistance robotisée d'insertion d'aiguille", Started in October 2016, Supervised by Alexandre Krupa (Lagadic, Inria) and Maud Marchal
- Rebecca Fribourg, "Perception and interaction with and via avatars", Started in September 2017, Supervised by Ferran Argelaguet, Ludovic Hoyet (Mimetic, Inria) and Anatole Lécuyer
- Guillaume Bataille, "Natural interactions with IoT using VR/AR", Started in October 2017, Supervised by Valérie Gouranton, Danielle Pelé (Orange Labs) and Bruno Arnaldi
- Xavier de Tinguy, "Haptic manipulation in virtual environments", Started in September 2017, Supervised by Maud Marchal, Claudio Pacchierotti (Lagadic, Inria) and Anatole Lécuyer
- Flavien Lecuyer, "Interactive digital introspection methods for archeology", Started in September 2017, Supervised by Valérie Gouranton, Grégor Marchand (CNRS) and Bruno Arnaldi
- Etienne Peillard, "Improving Perception and Interaction in Augmented Reality", Started in October 2017, Supervised by Guillaume Moreau, Ferran Argelaguet, Anatole Lécuyer and Jean-Marie Normand
- Antonin Bernardin, "Interactive physically-based simulation of dexterous manipulation for robot understanding", Started in September 2017, Supervised by Maud Marchal and Christian Duriez (Defrost, Inria)
- Romain Lagneau, "Data-driven models for dexterous manipulation of robots", Started in Septembre 2017, Supervised by Maud Marchal and Alexandre Krupa (Lagadic, Inria)
- Romain Terrier, "Presence of self and others in a collaborative virtual environment", Started in October 2017, Supervised by Valérie Gouranton, Thomas Boggini (b<>com) and Bruno Arnaldi
- Mathis Fleury, "Neurofeedback based on fMRI and EEG", Started in November 2017, Supervised by Anatole Lécuyer and Christian Barillot (Visages, Inria)

#### 10.2.2. Juries

#### 10.2.2.1. Selection committees

- Anatole Lécuyer was Member of selection committee of Assistant Professor Position at Ecole Centrale de Nantes.
- Valérie Gouranton was Member of selection committee of Assistant Professor Position at INSA Rennes, of Inria Junior Research Scientist (CR2 Inria) at Inria Rennes-Bretagne Atlantique
- Maud Marchal was Member of selection committee of Assistant Professor Position at Ecole Centrale de Nantes.
- Guillaume Moreau was Member of selection committee at Ecole Centrale de Nantes: Assistant Professor (2x), Full Professor (2x).

#### 10.2.2.2. PhD and HDR juries

- Anatole Lécuyer was Referee of PhD theses of Paul Issartel (Univ Paris Saclay), Cephise Louison (Univ Mediterrannée) and Joan Sol Roo (Univ Bordeaux).
- Bruno Arnaldi was Referee of PhD thesis of Damien Clergeaud (Univ Bordeaux).
- Maud Marchal was Referee of PhD thesis of Fanny Morin (Univ Grenoble Alpes).
- Guillaume Moreau was Referee of PhD thesis of Judicaël Menant (INSA Rennes), and Member of PhD committee of Abdelkader Bellarbi (Univ Paris-Saclay).
- Jean-Marie Normand was Member of the committee of Barteld Postma (Univ. Paris Saclay).

#### **10.3.** Popularization

The results of the team have been disseminated in various media coverages in 2017:

- "XENIUS" TV channel Arte (02/17) : presentation of Immersia activity in Cultural Heritage applications
- "Sciences et Avenir" magazine (02/17) : article on "BCI, VR and videogames"
- France-Info radio (07/17) : 4-min interview of Anatole Lécuyer on Virtual Reality
- 20-minutes magazine (10/17) : article on Immersia activity
- Virgin Radio (10/17) : interview of Ronan Gaugne on Immersia
- France 3 Bretagne (10/17) : presentation of "Journée Science et Musique" in Immersia
- France 3 Centre (11/17) : presentation of INTROSPECT project

The team has also participated to several dissemination events in 2017 (chronological order):

- "Ma Thèse en 180 secondes" (Rennes, 03/17) : Lorraine Perronnet presented her PhD.
- "Semaine du Cerveau 2017" (Rennes, 03/17) : presentation and demo of BCI by Nataliya Kosmyna.
- Pint of Science (05/17): presentation on virtual incarnation by Ferran Argelaguet.
- "Journées Science et Musique 2017" (Rennes, 10/17) : co-organization of this event, and presentation of several demos.
- "Journées du Pôle Aerospace Valley" (Agen, 12/17) : presentation by Guillaume Moreau.

## 11. Bibliography

#### **Publications of the year**

#### **Doctoral Dissertations and Habilitation Theses**

[1] L. PERRONNET. Combining EEG and FMRI for Neurofeedback, University of Rennes I, September 2017, https://hal.inria.fr/tel-01598667.

#### **Articles in International Peer-Reviewed Journal**

- [2] J.-B. BARREAU, R. GAUGNE, V. GOURANTON. *Immersive Point Cloud Manipulation for Cultural Heritage Documentation*, in "ERCIM News", October 2017, p. 1-3, https://hal.inria.fr/hal-01659814.
- [3] Y. GAFFARY, B. LE GOUIS, M. MARCHAL, F. ARGELAGUET, A. LÉCUYER, B. ARNALDI.AR Feels "Softer" than VR: Haptic Perception of Stiffness in Augmented versus Virtual Reality, in "IEEE Transactions on Visualization and Computer Graphics", November 2017, vol. 23, n<sup>o</sup> 11, p. 2372 - 2377 [DOI: 10.1109/TVCG.2017.2735078], https://hal.inria.fr/hal-01625290.
- [4] R. GAUGNE, F. NOUVIALE, O. RIOUAL, A. CHIRAT, K. GOHON, V. GOUPIL, M. TOUTIRAIS, B. BOSSIS, V. GOURANTON. EvoluSon: Walking through an Interactive History of Music, in "Presence: Teleoperators and Virtual Environments", December 2017, p. 1-23, https://hal.inria.fr/hal-01661727.
- [5] N. KOS'MYNA, A. LÉCUYER. Designing Guiding Systems for Brain-Computer Interfaces, in "Frontiers in Human Neuroscience", July 2017, vol. 11, p. 1-15 [DOI: 10.3389/FNHUM.2017.00396], https://hal.inria.fr/ hal-01669322.
- [6] J. LINDGREN. *As above, so below? Towards understanding inverse models in BCI*, in "Journal of Neural Engineering", December 2017, https://hal.inria.fr/hal-01669325.

- [7] T. NICOLAS, R. GAUGNE, V. GOURANTON. *Physical Digital Access Inside Archaeological Material*, in "ERCIM News", October 2017, vol. 111, p. 1-4, https://hal.inria.fr/hal-01659817.
- [8] L. PERRONNET, A. LÉCUYER, M. MANO, E. BANNIER, F. LOTTE, M. CLERC, C. BARILLOT. Unimodal Versus Bimodal EEG-fMRI Neurofeedback of a Motor Imagery Task, in "Frontiers in Human Neuroscience", April 2017, vol. 11 [DOI : 10.3389/FNHUM.2017.00193], https://hal.inria.fr/hal-01519755.

#### **International Conferences with Proceedings**

- [9] M. ACHIBET, B. LE GOUIS, M. MARCHAL, P.-A. LEZIART, F. ARGELAGUET SANZ, A. GIRARD, A. LÉCUYER, H. KAJIMOTO.*FlexiFingers: Multi-Finger Interaction in VR Combining Passive Haptics and Pseudo-Haptics*, in "IEEE Symposium on 3D User Interfaces", Los Angeles, United States, March 2017, https://hal.inria.fr/hal-01625154.
- [10] F. ARGELAGUET, M. DUCOFFE, A. LÉCUYER, R. GRIBONVAL. Spatial and rotation invariant 3D gesture recognition based on sparse representation, in "IEEE Symposium on 3D User Interfaces", Los Angeles, United States, March 2017, p. 158 - 167 [DOI: 10.1109/3DUI.2017.7893333], https://hal.inria.fr/hal-01625128.
- [11] J. BARBIER, P. KENNY, J. YOUNG, J.-M. NORMAND, M. T. KEANE, M. O'SULLIVAN, A. VEN-TRESQUE.MAAP Annotate: When Archaeology meets Augmented Reality for Annotation of Megalithic Art, in "VSMM 2017 - 23rd International Conference on Virtual Systems and Multimedia", Dublin, Ireland, October 2017, p. 1-8, https://hal.inria.fr/hal-01626057.
- [12] G. BOUYER, A. CHELLALI, A. LÉCUYER. Inducing self-motion sensations in driving simulators using forcefeedback and haptic motion, in "VR 2017 - 19th IEEE Virtual Reality", Los Angeles, United States, Proc. of the 19th IEEE Virtual Reality (VR 2017), IEEE, March 2017, p. 84-90 [DOI: 10.1109/VR.2017.7892234], https://hal.archives-ouvertes.fr/hal-01524537.
- [13] G. CORTES, E. MARCHAND, J. ARDOUIN, A. LÉCUYER. An Optical Tracking System based on Hybrid Stereo/Single-View Registration and Controlled Cameras, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'17", Vancouver, Canada, September 2017, p. 6185-6190, https://hal.inria.fr/hal-01562327.
- [14] G. CORTES, E. MARCHAND, J. ARDOUIN, A. LÉCUYER. Increasing Optical Tracking Workspace of VR Applications using Controlled Cameras, in "IEEE Symposium on 3D User Interfaces, 3DUI 2017", Los Angeles, United States, March 2017, p. 22-25 [DOI : 10.1109/3DUI.2017.7893313], https://hal.inria.fr/ hal-01446343.
- [15] A. EVAIN, F. ARGELAGUET, N. ROUSSEL, G. CASIEZ, A. LÉCUYER. Can I Think of Something Else when Using a BCI? Cognitive Demand of an SSVEP-based BCI, in "ACM Conference on Human Factors in Computing Systems", Denver, United States, May 2017, p. 5120-5125 [DOI: 10.1145/3025453.3026037], https://hal.inria.fr/hal-01625088.
- [16] B. LE GOUIS, F. LEHERICEY, M. MARCHAL, B. ARNALDI, V. GOURANTON, A. LÉCUYER. Haptic Rendering of FEM-based Tearing Simulation using Clusterized Collision Detection, in "WHC 2017 - IEEE World Haptics Conference", Munich, Germany, June 2017, p. 406-411, https://hal.inria.fr/hal-01675134.
- [17] B. LE GOUIS, M. MARCHAL, A. LÉCUYER, B. ARNALDI. *Elasticity-based Clustering for Haptic Interaction with Heterogeneous Deformable Objects*, in "VRIPHYS2017 13th Eurographics Workshop on Virtual Reality Interaction and Physical Simulation", Lyon, France, April 2017, p. 75-83, https://hal.inria.fr/hal-01675148.

- [18] G. LE MOULEC, F. ARGELAGUET, V. GOURANTON, A. BLOUIN, B. ARNALDI.AGENT: Automatic Generation of Experimental Protocol Runtime, in "ACM Symposium on Virtual Reality Software and Technology (VRST)", Gothenburg, Sweden, Virtual Reality Software and Technology, November 2017, https://hal. archives-ouvertes.fr/hal-01613873.
- [19] F. LOTTE, C. JEUNET. Online classification accuracy is a poor metric to study mental imagery-based bci user learning: an experimental demonstration and new metrics, in "7th International BCI Conference", Graz, Austria, September 2017, https://hal.archives-ouvertes.fr/hal-01519478.
- [20] M. MANO, E. BANNIER, L. PERRONNET, A. LÉCUYER, C. BARILLOT. *Hybrid EEG and fMRI platform for multi-modal neurofeedback*, in "International Society of Magnetic Resonance in Medicine", Honolulu, United States, Proc. Intl. Soc. Mag. Reson. Med. 25th, ISMRM, April 2017, 4550, http://www.hal.inserm.fr/inserm-01577442.
- [21] H. SI-MOHAMMED, F. ARGELAGUET, G. CASIEZ, N. ROUSSEL, A. LÉCUYER. Brain-Computer Interfaces and Augmented Reality: A State of the Art, in "Graz Brain-Computer Interface Conference", Graz, Austria, September 2017 [DOI: 10.3217/978-3-85125-533-1-82], https://hal.inria.fr/hal-01625167.
- [22] A. TRANSON, A. VERHULST, J.-M. NORMAND, G. MOREAU, M. SUGIMOTO. Evaluation of Facial Expressions as an Interaction Mechanism and their Impact on Affect, Workload and Usability in an AR game, in "VSMM 2017 - 23rd International Conference on Virtual Systems and Multimedia", Dublin, Ireland, October 2017, p. 1-8, https://hal.archives-ouvertes.fr/hal-01625955.
- [23] A. VERHULST, J.-M. NORMAND, C. LOMBART, G. MOREAU.A study on the use of an immersive Virtual Reality store to investigate consumer perceptions and purchase behavior toward non-standard fruits and vegetables, in "IEEE Virtual Reality", Los Angeles, United States, March 2017, p. 55 - 63 [DOI: 10.1109/VR.2017.7892231], https://hal.archives-ouvertes.fr/hal-01619314.

#### **Conferences without Proceedings**

- [24] J.-B. BARREAU, A. GAGNIER, R. GAUGNE, G. MARCHAND, J. CALVO GÓMEZ, R. COLLETER, V. GOURANTON. *A multi-modal digital analysis of a mesolithic clavicle : preserving and studying the oldest human bone in Brittany*, in "2017 3D Imaging in Cultural Heritage Conference", London, United Kingdom, November 2017, https://hal.inria.fr/hal-01659855.
- [25] J.-B. BARREAU, E. LANOË, R. GAUGNE.3D reconstruction of the fortified entrance of the Citadel of Aleppo from a few sightseeing photos, in "DCH 2017 - Interdisciplinary Conference on Digital Cultural Heritage", Berlin, Germany, August 2017, p. 1-2, https://hal.inria.fr/hal-01659865.
- [26] F. COUSSEAU, Y. BERNARD, Q. PETIT, J.-B. BARREAU, R. GAUGNE, L. QUESNEL. Relever les élévations d'une architecture mégalithique : développements méthodologiques, in "GMPCA 2017 - XXIème Colloque international du Groupe des Méthodes Pluridisciplinaires Contribuant à l'Archéologie", Rennes, France, April 2017, https://hal.inria.fr/hal-01659868.
- [27] R. GAUGNE, T. NICOLAS, J.-B. BARREAU, G. MARCHAND, R. AUGER, P. FRANCUS, V. GOURAN-TON.*Méthodes d'introspection numérique pour les objets archéologiques*, in "GMPCA 2017 - XXIème colloque international du Groupe des Méthodes Pluridisciplinaires Contribuant à l'Archéologie", rennes, France, April 2017, https://hal.inria.fr/hal-01659861.

- [28] T. NICOLAS, R. GAUGNE, C. TAVERNIER, É. MILLET, B. R. ROSEAU, R. BERNADET, B. ARNALDI, V. GOURANTON. Lift the veil off the block samples from the Warcq chariot burial: CT-scan, photogrammetry, 3D printing and mixed reality, in "3D imaging in Cultural Heritage", London, United Kingdom, November 2017, https://hal.inria.fr/hal-01661232.
- [29] A. VERHULST, J.-M. NORMAND, G. MOREAU. Generation of variability in shape, aspect and time of 3D Fruits and Vegetables, in "VSMM 2017 - 23rd International Conference on Virtual Systems and Multimedia", Dublin, Ireland, October 2017, p. 1-8, https://hal.archives-ouvertes.fr/hal-01625956.

#### Scientific Books (or Scientific Book chapters)

[30] C. JEUNET, S. DEBENER, F. LOTTE, J. MATTOUT, R. SCHERER, C. ZICH.*Mind the Traps! Design Guidelines for Rigorous BCI Experiments*, in "Brain–Computer Interfaces Handbook: Technological and Theoretical Advances", C. S. NAM, A. NIJHOLT, F. LOTTE (editors), CRC Press, 2018, https://hal.inria.fr/hal-01620186.

#### **Other Publications**

- [31] G. CORTES, F. ARGELAGUET SANZ, E. MARCHAND, A. LÉCUYER. Toward Application-Driven VR Systems: Analysis of Head and Hand 3D Motions in a Specific CAVE-based Application, March 2017, IEEE Symposium on 3D User Interfaces, 3DUI 2017, poster session, Poster, https://hal.inria.fr/hal-01482150.
- [32] G. LE MOULEC, A. BLOUIN, V. GOURANTON, B. ARNALDI. *Automatic Production of End User Documentation for DSLs*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01549042.

#### **References in notes**

- [33] D. A. BOWMAN, E. KRUIJFF, J. J. LAVIOLA, I. POUPYREV. 3D User Interfaces: Theory and Practice, Addison Wesley, 2004.
- [34] A. LÉCUYER.Simulating Haptic Feedback Using Vision: A Survey of Research and Applications of Pseudo-Haptic Feedback, in "Presence: Teleoperators and Virtual Environments", January 2009, vol. 18, n<sup>o</sup> 1, p. 39–53, http://www.mitpressjournals.org/doi/abs/10.1162/pres.18.1.39.

# **Project-Team HYCOMES**

# Modélisation hybride & conception par contrats pour les systèmes embarqués multi-physiques

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: Université Rennes 1

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Embedded and Real-time Systems

## **Table of contents**

1.	Personnel	410
2.	Overall Objectives	410
3.	Research Program	411
	3.1. Hybrid Systems Modeling	411
	3.2. Background on non-standard analysis	411
	3.3. Contract-Based Design, Interfaces Theories, and Requirements Engineering	412
4.	New Software and Platforms	414
	4.1. Demodocos	414
	4.2. MICA	415
	4.3. TnF-C++	415
5.	New Results	415
	5.1. Semantics, Static or Runtime Analysis of Hybrid Systems	415
	5.1.1. Structural Analysis of Multi-Mode DAEs	415
	5.1.2. Operational Models for Piecewise-Smooth Systems	415
	5.1.3. Accelerated Simulation of Hybrid Systems: Method combining static analysis and ru	untime
	execution analysis	416
	5.1.4. A Type-based Analysis of Causality Loops in Hybrid Systems Modelers	416
	5.2. Formal Verification of Hybrid Systems	416
	5.2.1. Formal Verification of Station Keeping Maneuvers for a Planar Autonomous I	Hybrid
	System	416
	5.2.2. Formal verification of obstacle avoidance and navigation of ground robots	416
	5.3. Synchronous Interfaces and Assume/Guarantee Contracts	417
	5.4. CominWeb project of the Labex CominLabs	417
6.	Bilateral Contracts and Grants with Industry	418
7.	Partnerships and Cooperations	418
	7.1. Regional Initiatives	418
	7.2. National Initiatives	419
8.	Dissemination	419
	8.1. Promoting Scientific Activities	419
	8.1.1. Scientific Events Organisation	419
	8.1.2. Scientific Events Selection	419
	8.1.2.1. Member of the Conference Program Committees	419
	8.1.2.2. Reviewer	420
	8.1.3. Journal	420
	8.1.4. Invited Talks	420
	8.1.5. Scientific Expertise	420
	8.1.6. Research Administration	420
	8.2. Teaching - Supervision - Juries	420
	8.2.1. Teaching	420
	8.2.2. Supervision	420
	8.2.3. Juries	420
9.	Bibliography	421

### **Project-Team HYCOMES**

*Creation of the Team: 2013 July 01, updated into Project-Team: 2016 September 01* **Keywords:** 

#### **Computer Science and Digital Science:**

- A2. Software
- A2.1. Programming Languages
- A2.1.1. Semantics of programming languages
- A2.1.5. Constraint programming
- A2.1.8. Synchronous languages
- A2.1.10. Domain-specific languages
- A2.2. Compilation
- A2.3. Embedded and cyber-physical systems
- A2.3.1. Embedded systems
- A2.3.2. Cyber-physical systems
- A2.3.3. Real-time systems
- A2.4. Verification, reliability, certification
- A2.4.1. Analysis
- A2.4.2. Model-checking
- A2.4.3. Proofs
- A2.5. Software engineering
- A2.5.1. Software Architecture & Design
- A2.5.2. Component-based Design
- A3. Data and knowledge
- A3.1. Data
- A3.1.1. Modeling, representation
- A6. Modeling, simulation and control
- A6.1. Mathematical Modeling
- A6.1.1. Continuous Modeling (PDE, ODE)
- A6.1.3. Discrete Modeling (multi-agent, people centered)
- A6.1.5. Multiphysics modeling
- A8.4. Computer Algebra

#### **Other Research Topics and Application Domains:**

- B2. Health
- B2.4. Therapies
- B2.4.3. Surgery
- B4. Energy
- B4.4. Energy delivery
- B4.4.1. Smart grids
- B5. Industry of the future
- B5.2. Design and manufacturing
- B5.2.1. Road vehicles

B5.2.2. - Railway
B5.2.3. - Aviation
B5.2.4. - Aerospace
B5.8. - Learning and training
B5.9. - Industrial maintenance
B7. - Transport and logistics
B7.1. - Traffic management
B7.1.3. - Air traffic
B8. - Smart Cities and Territories
B8.1. - Smart building/home
B8.1.1. - Energy for smart buildings

## 1. Personnel

#### **Research Scientists**

Benoît Caillaud [Team leader, Inria, Senior Researcher, HDR] Albert Benveniste [Inria, Emeritus, HDR] Khalil Ghorbal [Inria, Researcher]

#### **Technical Staff**

Jean Hany [Inria, from Dec 2017] Aurélien Lamercerie [Inria, until Oct 2017]

#### **PhD Students**

Ayman Aljarbouh [Inria, graduated Sep 2017, currently postdoctoral researcher at Verimag. Grenoble, France] Aurélien Lamercerie [Univ de Rennes I, from Nov 2017]

#### Administrative Assistants

Angélique Jarnoux [Inria] Armelle Mozziconacci [CNRS]

## 2. Overall Objectives

#### 2.1. Overall Objectives

Hycomes is a team of the Rennes — Bretagne Atlantique Inria research center since July 2013. The team builds upon the most promising results of the former S4 team-project and of the Synchronics large scale initiative. Two topics in cyber-physical systems design are covered:

- Hybrid systems modelling, with an emphasis on the design of modelling languages in which software systems, in interaction with a complex physical environment, can be modelled, simulated and verified. A special attention is paid to the mathematical rigorous semantics of these languages, and to the correctness (wrt. such semantics) of the simulations and of the static analyses that must be performed during compilation. The Modelica language is the main application field. The team aims at contributing language extensions facilitating the modelling of physical domains which are poorly supported by the Modelica language. The Hycomes team is also designing new structural analysis methods for hybrid (aka. multi-mode) Modelica models. New simulation and verification techniques for large Modelica models are also in the scope of the team.
- Contract-based design and interface theories, with applications to requirements engineering in the context of safety-critical systems design.

## 3. Research Program

#### 3.1. Hybrid Systems Modeling

Systems industries today make extensive use of mathematical modeling tools to design computer controlled physical systems. This class of tools addresses the modeling of physical systems with models that are simpler than usual scientific computing problems by using only Ordinary Differential Equations (ODE) and Difference Equations but not Partial Differential Equations (PDE). This family of tools first emerged in the 1980's with SystemBuild by MatrixX (now distributed by National Instruments) followed soon by Simulink by Mathworks, with an impressive subsequent development.

In the early 90's control scientists from the University of Lund (Sweden) realized that the above approach did not support component based modeling of physical systems with reuse <sup>0</sup>. For instance, it was not easy to draw an electrical or hydraulic circuit by assembling component models of the various devices. The development of the Omola language by Hilding Elmqvist was a first attempt to bridge this gap by supporting some form of Differential Algebraic Equations (DAE) in the models. Modelica quickly emerged from this first attempt and became in the 2000's a major international concerted effort with the Modelica Consortium <sup>0</sup>. A wider set of tools, both industrial and academic, now exists in this segment <sup>0</sup>. In the EDA sector, VHDL-AMS was developed as a standard [13].

Despite these tools are now widely used by a number of engineers, they raise a number of technical difficulties. The meaning of some programs, their mathematical semantics, can be tainted with uncertainty. A main source of difficulty lies in the failure to properly handle the discrete and the continuous parts of systems, and their interaction. How the propagation of mode changes and resets should be handled? How to avoid artifacts due to the use of a global ODE solver causing unwanted coupling between seemingly non interacting subsystems? Also, the mixed use of an equational style for the continuous dynamics with an imperative style for the mode changes and resets is a source of difficulty when handling parallel composition. It is therefore not uncommon that tools return complex warnings for programs with many different suggested hints for fixing them. Yet, these "pathological" programs can still be executed, if wanted so, giving surprising results — See for instance the Simulink examples in [20], [1] and [16].

Indeed this area suffers from the same difficulties that led to the development of the theory of synchronous languages as an effort to fix obscure compilation schemes for discrete time equation based languages in the 1980's. Our vision is that hybrid systems modeling tools deserve similar efforts in theory as synchronous languages did for the programming of embedded systems.

#### 3.2. Background on non-standard analysis

Non-Standard analysis plays a central role in our research on hybrid systems modeling [1], [20], [17], [16]. The following text provides a brief summary of this theory and gives some hints on its usefulness in the context of hybrid systems modeling. This presentation is based on our paper [1], a chapter of Simon Bliudze's PhD thesis [25], and a recent presentation of non-standard analysis, not axiomatic in style, due to the mathematician Lindström [47].

Non-standard numbers allowed us to reconsider the semantics of hybrid systems and propose a radical alternative to the *super-dense time semantics* developed by Edward Lee and his team as part of the Ptolemy II project, where cascades of successive instants can occur in zero time by using  $\mathbb{R}_+ \times \mathbb{N}$  as a time index. In the non-standard semantics, the time index is defined as a set  $\mathbb{T} = \{n\partial \mid n \in *\mathbb{N}\}$ , where  $\partial$  is an *infinitesimal* and  $*\mathbb{N}$  is the set of *non-standard integers*. Remark that (1)  $\mathbb{T}$  is dense in  $\mathbb{R}_+$ , making it "continuous", and (2) every  $t \in \mathbb{T}$  has a predecessor in  $\mathbb{T}$  and a successor in  $\mathbb{T}$ , making it "discrete". Although it is not effective from a computability point of view, the *non-standard semantics* provides a framework that is familiar to the

<sup>&</sup>lt;sup>0</sup>http://www.lccc.lth.se/media/LCCC2012/WorkshopSeptember/slides/Astrom.pdf

<sup>&</sup>lt;sup>0</sup>https://www.modelica.org/

<sup>&</sup>lt;sup>0</sup>SimScape by Mathworks, Amesim by LMS International, now Siemens PLM, and more.

computer scientist and at the same time efficient as a symbolic abstraction. This makes it an excellent candidate for the development of provably correct compilation schemes and type systems for hybrid systems modeling languages.

Non-standard analysis was proposed by Abraham Robinson in the 1960s to allow the explicit manipulation of "infinitesimals" in analysis [53], [41], [12]. Robinson's approach is axiomatic; he proposes adding three new axioms to the basic Zermelo-Fraenkel (ZFC) framework. There has been much debate in the mathematical community as to whether it is worth considering non-standard analysis instead of staying with the traditional one. We do not enter this debate. The important thing for us is that non-standard analysis allows the use of the non-standard discretization of continuous dynamics "as if" it was operational.

Not surprisingly, such an idea is quite ancient. Iwasaki et al. [43] first proposed using non-standard analysis to discuss the nature of time in hybrid systems. Bliudze and Krob [26], [25] have also used non-standard analysis as a mathematical support for defining a system theory for hybrid systems. They discuss in detail the notion of "system" and investigate computability issues. The formalization they propose closely follows that of Turing machines, with a memory tape and a control mechanism.

## **3.3. Contract-Based Design, Interfaces Theories, and Requirements** Engineering

System companies such as automotive and aeronautic companies are facing significant difficulties due to the exponentially raising complexity of their products coupled with increasingly tight demands on functionality, correctness, and time-to-market. The cost of being late to market or of imperfections in the products is staggering as witnessed by the recent recalls and delivery delays that many major car and airplane manufacturers had to bear in the recent years. The specific root causes of these design problems are complex and relate to a number of issues ranging from design processes and relationships with different departments of the same company and with suppliers, to incomplete requirement specification and testing.

We believe the most promising means to address the challenges in systems engineering is to employ structured and formal design methodologies that seamlessly and coherently combine the various viewpoints of the design space (behavior, space, time, energy, reliability, ...), that provide the appropriate abstractions to manage the inherent complexity, and that can provide correct-by-construction implementations. The following technology issues must be addressed when developing new approaches to the design of complex systems:

- The overall design flows for heterogeneous systems and the associated use of models across traditional boundaries are not well developed and understood. Relationships between different teams inside a same company, or between different stake-holders in the supplier chain, are not well supported by solid technical descriptions for the mutual obligations.
- System requirements capture and analysis is in large part a heuristic process, where the informal text and natural language-based techniques in use today are facing significant challenges. Formal requirements engineering is in its infancy: mathematical models, formal analysis techniques and links to system implementation must be developed.
- Dealing with variability, uncertainty, and life-cycle issues, such as extensibility of a product family, are not well-addressed using available systems engineering methodologies and tools.

The challenge is to address the entire process and not to consider only local solutions of methodology, tools, and models that ease part of the design.

*Contract-based design* has been proposed as a new approach to the system design problem that is rigorous and effective in dealing with the problems and challenges described before, and that, at the same time, does not require a radical change in the way industrial designers carry out their task as it cuts across design flows of different type. Indeed, contracts can be used almost everywhere and at nearly all stages of system design, from early requirements capture, to embedded computing infrastructure and detailed design involving circuits and other hardware. Contracts explicitly handle pairs of properties, respectively representing the assumptions on the environment and the guarantees of the system under these assumptions. Intuitively, a contract is a pair

C = (A, G) of assumptions and guarantees characterizing in a formal way 1) under which context the design is assumed to operate, and 2) what its obligations are. Assume/Guarantee reasoning has been known for a long time, and has been used mostly as verification mean for the design of software [51]. However, contract based design with explicit assumptions is a philosophy that should be followed all along the design, with all kinds of models, whenever necessary. Here, specifications are not limited to profiles, types, or taxonomy of data, but also describe the functions, performances of various kinds (time and energy), and reliability. This amounts to enrich a component's interface with, on one hand, formal specifications of the behavior of the environment in which the component may be instantiated and, on the other hand, of the expected behavior of the component itself. The consideration of rich interfaces is still in its infancy. So far, academic researchers have addressed the mathematics and algorithmics of interfaces theories and contract-based reasoning. To make them a technique of choice for system engineers, we must develop:

- Mathematical foundations for interfaces and requirements engineering that enable the design of frameworks and tools;
- A system engineering framework and associated methodologies and tool sets that focus on system requirements modeling, contract specification, and verification at multiple abstraction layers.

A detailed bibliography on contract and interface theories for embedded system design can be found in [2]. In a nutshell, contract and interface theories fall into two main categories:

- Assume/guarantee contracts. By explicitly relying on the notions of assumptions and guarantees, A/Gcontracts are intuitive, which makes them appealing for the engineer. In A/G-contracts, assumptions and guarantees are just properties regarding the behavior of a component and of its environment. The typical case is when these properties are formal languages or sets of traces, which includes the class of safety properties [44], [33], [50], [15], [34]. Contract theories were initially developed as specification formalisms able to refuse some inputs from the environment [42]. A/G-contracts were advocated by the SPEEDS project [19]. They were further experimented in the framework of the CESAR project [37], with the additional consideration of *weak* and *strong* assumptions. This is still a very active research topic, with several recent contributions dealing with the timed [24] and probabilistic [29], [30] viewpoints in system design, and even mixed-analog circuit design [52].
- Automata theoretic interfaces. Interfaces combine assumptions and guarantees in a single, automata theoretic specification. Most interface theories are based on Lynch Input/Output Automata [49], [48]. Interface Automata [56], [55], [57], [31] focus primarily on parallel composition and compatibility: Two interfaces can be composed and are compatible if there is at least one environment where they can work together. The idea is that the resulting composition exposes as an interface the needed information to ensure that incompatible pairs of states cannot be reached. This can be achieved by using the possibility, for an Interface Automaton, to refuse selected inputs from the environment in a given state, which amounts to the implicit assumption that the environment will never produce any of the refused inputs, when the interface is in this state. Modal Interfaces [3] inherit from both Interface Automata and the originally unrelated notion of Modal Transition System [46], [14], [27], [45]. Modal Interfaces are strictly more expressive than Interface Automata by decoupling the I/O orientation of an event and its deontic modalities (mandatory, allowed or forbidden). Informally, a must transition is available in every component that realizes the modal interface, while a may transition needs not be. Research on interface theories is still very active. For instance, timed [58], [21], [23], [39], [38], [22], probabilistic [29], [40] and energy-aware [32] interface theories have been proposed recently.

Requirements Engineering is one of the major concerns in large systems industries today, particularly so in sectors where certification prevails [54]. DOORS projects collecting requirements are poorly structured and cannot be considered a formal modeling framework today. They are nothing more than an informal documentation enriched with hyperlinks. As examples, medium size sub-systems may have a few thousands requirements and the Rafale fighter aircraft has above 250,000 of them. For the Boeing 787, requirements were not stable while subcontractors performed the development of the fly-by-wire and of the landing gear subsystems.

We see Contract-Based Design and Interfaces Theories as innovative tools in support of Requirements Engineering. The Software Engineering community has extensively covered several aspects of Requirements Engineering, in particular:

- the development and use of large and rich ontologies; and
- the use of Model Driven Engineering technology for the structural aspects of requirements and resulting hyperlinks (to tests, documentation, PLM, architecture, and so on).

Behavioral models and properties, however, are not properly encompassed by the above approaches. This is the cause of a remaining gap between this phase of systems design and later phases where formal model based methods involving behavior have become prevalent—see the success of Matlab/Simulink/Scade technologies. We believe that our work on contract based design and interface theories is best suited to bridge this gap.

## 4. New Software and Platforms

#### 4.1. Demodocos

Demodocos (Examples to Generic Scenario Models Generator)

KEYWORDS: Surgical process modelling - Net synthesis - Process mining

SCIENTIFIC DESCRIPTION: Demodocos is used to construct a Test and Flip net (Petri net variant) from a collection of instances of a given procedure. The tool takes as input either standard XES log files (a standard XML file format for process mining tools) or a specific XML file format for surgical applications. The result is a Test and Flip net and its marking graph. The tool can also build a #SEVEN scenario for integration into a virtual reality environment. The scenario obtained corresponds to the generalization of the input instances, namely the instances synthesis enriched with new behaviors respecting the relations of causality, conflicts and competition observed.

Demodocos is a synthesis tool implementing a linear algebraic polynomial time algorithm. Computations are done in the Z/2Z ring. Test and Flip nets extend Elementary Net Systems by allowing test to zero, test to one and flip arcs. The effect of flip arcs is to complement the marking of the place. While the net synthesis problem has been proved to be NP hard for Elementary Net Systems, thanks to flip arcs, the synthesis of Test and Flip nets can be done in polynomial time. Test and flip nets have the required expressivity to give concise and accurate representations of surgical processes (models of types of surgical operations). Test and Flip nets can express causality and conflict relations. The tool takes as input either standard XES log files (a standard XML file format for process mining tools) or a specific XML file format for surgical applications. The output is a Test and Flip net, solution of the following synthesis problem: Given a finite input language (log file), compute a net, which language is the least language in the class of Test and Flip net languages, containing the input language.

FUNCTIONAL DESCRIPTION: The tool Demodocos allows to build a generic model for a given procedure from some examples of instances of this procedure. The generated model can take the form of a graph, a Test 'n Flip net or a SEVEN scenario (intended for integration into a virtual reality environment).

The classic use of the tool is to apply the summary operation to a set of files describing instances of the target procedure. Several file formats are supported, including the standard XES format for log events. As output, several files are generated. These files represent the generic procedure in different forms, responding to varied uses.

This application is of limited interest in the case of an isolated use, out of context and without a specific objective when using the model generated. It was developed as part of a research project focusing in particular on surgical procedures, and requiring the generation of a generic model for integration into a virtual reality training environment. It is also quite possible to apply the same method in another context.

- Participants: Aurélien Lamercerie and Benoît Caillaud
- Contact: Benoît Caillaud
- Publication: Surgical Process Mining with Test and Flip Net Synthesis
- URL: http://tinyurl.com/oql6f3y

#### **4.2. MICA**

Model Interface Compositional Analysis Library

KEYWORDS: Modal interfaces - Contract-based desing

SCIENTIFIC DESCRIPTION: In Mica, systems and interfaces are represented by extension. However, a careful design of the state and event heap enables the definition, composition and analysis of reasonably large systems and interfaces. The heap stores states and events in a hash table and ensures structural equality (there is no duplication). Therefore complex data-structures for states and events induce a very low overhead, as checking equality is done in constant time.

Thanks to the Inter module and the mica interactive environment, users can define complex systems and interfaces using Ocaml syntax. It is even possible to define parameterized components as Ocaml functions. FUNCTIONAL DESCRIPTION: Mica is an Ocaml library implementing the Modal Interface algebra. The purpose of Modal Interfaces is to provide a formal support to contract based design methods in the field of system engineering. Modal Interfaces enable compositional reasoning methods on I/O reactive systems.

- Participant: Benoît Caillaud
- Contact: Benoît Caillaud
- URL: http://www.irisa.fr/s4/tools/mica/

#### 4.3. TnF-C++

FUNCTIONAL DESCRIPTION: TnF-C++ is a robust and portable re-implementation of Flipflop, developed in 2014 and integrated in the S3PM toolchain. Both software have been designed in the context of the S3PM project on surgical procedure modeling and simulation,

• Contact: Benoît Caillaud

## 5. New Results

#### 5.1. Semantics, Static or Runtime Analysis of Hybrid Systems

#### 5.1.1. Structural Analysis of Multi-Mode DAEs

Differential Algebraic Equation (DAE) systems constitute the mathematical model supporting physical modeling languages such as Modelica or Simscape. Unlike Ordinary Differential Equations, or ODEs, they exhibit subtle issues because of their implicit *latent equations* and related *differentiation index*. Multi-mode DAE (mDAE) systems are much harder to deal with, not only because of their mode-dependent dynamics, but essentially because of the events and resets occurring at mode transitions. Unfortunately, the large literature devoted to the numerical analysis of DAEs do not cover the multi-mode case. It typically says nothing about mode changes. This lack of foundations cause numerous difficulties to the existing modeling tools. Some models are well handled, others are not, with no clear boundary between the two classes. In [11], we develop a comprehensive mathematical approach to the *structural analysis* of mDAE systems which properly extends the usual analysis of DAE systems. We define a constructive semantics based on nonstandard analysis and show how to produce execution schemes in a systematic way. This work has been accepted for presentation at the HSCC 2017 conference [18] in April 2017.

#### 5.1.2. Operational Models for Piecewise-Smooth Systems

In [7], we study ways of constructing meaningful operational models of piecewise-smooth systems (PWS). The systems we consider are described by polynomial vector fields defined on non-overlapping semi-algebraic sets, which form a partition of the state space. Our approach is to give meaning to motion in systems of this type by automatically synthesizing operational models in the form of hybrid automata (HA). Despite appearances, it is in practice often difficult to arrive at satisfactory HA models of PWS. The different ways of building operational models that we explore in our approach can be thought of as defining different semantics for the underlying PWS. These differences have a number of interesting nuances related to phenomena such as chattering, non-determinism, so-called mythical modes and sliding behaviour.

## 5.1.3. Accelerated Simulation of Hybrid Systems: Method combining static analysis and runtime execution analysis

Ayman Aljarbouh has defended his PhD [4] on September 13th 2017. His PhD has been partially funded by an ARED grant of the Brittany Regional Council. His doctoral work took place in the context of the Modrio (completed in 2016) and Sys2Soft (completed in 2015) projects on hybrid systems modeling. Ayman Aljarbouh has been working on accelerated simulation techniques for hybrid systems. In particular, he has contributed, and implemented in a software prototype, a regularisation method transforming automatically at runtime a chattering behaviour into a semantics preserving smooth behaviour. He has also contributed a method for the approximation of Zeno behaviour. This method enables to jump past an accumulation of an infinite number of zero-crossing events, and to continue the simulation of a large class of Zeno hybrid systems, after accumulation points.

#### 5.1.4. A Type-based Analysis of Causality Loops in Hybrid Systems Modelers

Explicit hybrid systems modelers like Simulink/Stateflow allow for programming both discrete- and continuous-time behaviors with complex interactions between them. A key issue in their compilation is the static detection of algebraic or causality loops. Such loops can cause simulations to deadlock and prevent the generation of statically scheduled code. In [5], we addresses this issue for a hybrid modeling language that combines synchronous data-flow equations with Ordinary Differential Equations (ODEs). We introduce the operator last(x) for the left-limit of a signal x. This operator is used to break causality loops and permits a uniform treatment of discrete and continuous state variables. The semantics relies on non-standard analysis, defining an execution as a sequence of infinitesimally small steps. A signal is deemed causally correct when it can be computed sequentially and only changes infinitesimally outside of announced discrete events like zero-crossings. The causality analysis takes the form of a type system that expresses dependences between signals. In well-typed programs, signals are provably continuous during integration provided that imported external functions are also continuous. The effectiveness of this system is illustrated with several examples written in Zélus, a Lustre-like synchronous language extended with hierarchical automata and ODEs.

#### 5.2. Formal Verification of Hybrid Systems

#### 5.2.1. Formal Verification of Station Keeping Maneuvers for a Planar Autonomous Hybrid System

In [9], we investigate the formal verification of a hybrid control law designed to perform a station keeping maneuver for a planar vehicle. Such maneuver requires that the vehicle reaches a neighborhood of its station in finite time and remains in it while waiting for further commands. We model the dynamics as well as the control law as a hybrid program and formally verify the reachability and safety properties involved. We highlight in particular the automated generation of invariant regions which turns out to be crucial in performing such verification. We use the hybrid system theorem prover KeymaeraX to formally check the parts of the proof that can be automatized in the current state of the tool.

#### 5.2.2. Formal verification of obstacle avoidance and navigation of ground robots

In [6], we answer fundamental safety questions for ground robot navigation: Under which circumstances does a given control decision make a ground robot safely avoid obstacles? Unsurprisingly, the answer depends on the exact formulation of the safety objective as well as the physical capabilities and limitations of the robot and the obstacles. Because uncertainties about the exact future behavior of a robot's environment make this a challenging problem, we formally verify corresponding controllers and provide rigorous safety proofs justifying why they can never collide with the obstacle in the respective physical model. To account for ground robots in which different physical phenomena are important, we analyze a series of increasingly strong properties of controllers for increasingly rich dynamics and identify the impact that the additional model parameters have on the required safety margins. We analyze and formally verify: (i) static safety, which ensures that no collisions can happen with stationary obstacles, (ii) passive safety, which ensures that no collisions can happen with stationary obstacles while the robot moves, (iii) the stronger passive

friendly safety in which the robot further maintains sufficient maneuvering distance for obstacles to avoid collision as well, and (iv) passive orientation safety, which allows for imperfect sensor coverage of the robot, i. e., the robot is aware that not everything in its environment will be visible. We formally prove that safety can be guaranteed despite sensor uncertainty and actuator perturbation. We complement these provably correct safety properties with liveness properties: we prove that provably safe motion is flexible enough to let the robot navigate waypoints and pass intersections. In order to account for the mixed influence of discrete control decisions and the continuous physical motion of the ground robot, we develop corresponding hybrid system models and use differential dynamic logic theorem proving techniques to formally verify their correctness. Since these models identify a broad range of conditions under which control decisions are provably safe, our results apply to any control algorithm for ground robots with the same dynamics. As a demonstration, we, thus, also synthesize provably correct runtime monitor conditions that check the compliance of any control algorithm with the verified control decisions.

#### 5.3. Synchronous Interfaces and Assume/Guarantee Contracts

In [10], we establish a link between the theory of Moore Interfaces proposed in 2002 by Chakraborty et al. as a specification framework for synchronous transition systems, and the Assume/Guarantee contracts as proposed in 2007 by Benveniste et al. as a simple and flexible contract framework. As our main result we show that the operation of saturation of A/G contracts (namely the mapping  $(A, G) \rightarrow (A, G \lor \neg A)$ ), which was considered a drawback of this theory, is indeed implemented by the Moore Game of Chakraborty et al. We further develop this link and come up with some remarks on Moore Interfaces.

#### 5.4. CominWeb project of the Labex CominLabs

Jean Hany and Albert Benveniste (together with William Dedzoe) were involved in this project.

CominWeb is a project supported by the Labex CominLabs since 2013. Its original objective was to equip CominLabs with Web infrastructures, tools, and services, that would allow to run the scientific activity of the Labex in an innovative way. Based on a study of the population of the CominLabs researchers, performed in year 2014-15 by the teams of CominLabs involved in social sciences, several services were investigated and prototyped. A short trial addressed the automatic generation of a scientific activity report, for a CominLabs project, from the material available from the publications ot the project team. This was suspended because such a service was not considered very useful by the community. A second trial (nicknamed "NSA") consisted in monitoring the flows of email exchanges addressed to aliases of the CominLabs projects, with the objective of classifying the mails into: meeting announcements, mails with attachments of interest, and other mails. This would give to the CominLabs head a view on the project's activities without asking for any specific contribution from the researchers. This was more interesting. Still, a difficulty was that researchers did not use the project aliases so much. For priority issues, this development was also suspended.

The main result of this project is thus the service called *LookinLabs*, deployed in two different versions: http:// lookinlabs4halinria.cominlabs.ueb.eu/ and http://www.lookinlabs.cominlabs.ueb.eu/. The former is a more advanced version of LookinLabs, developed for the whole Inria community, by exploiting the HAL publication archive. LookinLabs for HAL-Inria allows the user to find, among teams/individuals/publications taken from all the Inria teams, those best matching a query consisting of a list of keywords or a short text. The tool exploits, as data, HAL-Inria archives, in combination with the Inria Activity reports (the Raweb), and the internal data base of Inria teams called BASTRI. Active teams/individuals are shown in boldface. Teams/individuals shown in gray are no longer active at Inria. If team TEAM0 is no longer active, the mention: TEAM0  $\rightarrow$ (TEAM1,TEAM2) indicates follow-up active teams, if any. In LookinLabs, no ontology is used. No data need to be manually entered (besides the users' queries). The tool uses *Elasticsearch* (https://www.elastic.co/fr/ products/elasticsearch) as its core algorithm. This means that the matching is based on a distance between the query and the set of data attached, in HAL, to each team/individual/publication. Ranking is performed accordingly. Explanations are given for each returned item. Correlation graphs are given, allowing to navigate through teams or individuals that share common interests (they may or may not be co-authors). LookinLabs is deployed in two versions. LookinLabs4HALInria is the one we just described. The other version is in operation since 2016 and addresses the scientific community of CominLabs researchers. The data used are up to 10 standard bibliographical data bases (Dblp, IEEE Explore, Arxiv, HAL, and more) for which links have been collected from the researchers (this was the only data they were asked for). Results are returned in the form of individuals and publications, not teams.

## 6. Bilateral Contracts and Grants with Industry

#### 6.1. GLOSE

The simulation of system-level models requires synchronizing, at simulation-time, physical models with software models. These models are developed and maintained by different stakeholders: physics engineers, control engineers and software engineers. Models designed by physics engineers are either detailed 3D finite-elements models, with partial differential equations (PDEs), or finite-dimension 0D models (obtained by model reduction techniques, or by empirical knowledge) expressed in modeling languages such as Simulink (with ordinary differential equations, or ODEs), Modelica (with differential algebraic equations, or DAEs), or directly as a C code embedding both the differential equations and its discretization scheme. Control engineers favor Matlab/Simulink, mainly because of its toolboxes and ease of use. Computer scientists program or model real-time reactive software, either with a dedicated language, for instance SCADE, hierarchical state machines or sequence/activity diagrams (as in UML/SysML) or directly in C. Coupling together heterogeneous models and programs, so that they can be co-simulated, is not only a technological challenge, but more importantly raises several deep and difficult questions: Can we trust simulations? What about their reproducibility? Will it be possible to simulate large systems with hundreds to thousands of component models?

The objective of the GLOSE project is to address these objectives, and propose both sound foundations and practical technological solutions to system level modeling and simulation. The GLOSE project has started in December 2017 and is funded by Safran, in the realm of the DESIR joint Safran-Academia research network. The academic teams contributing to GLOSE are the Hycomes, Diverse and Kairos Inria teams, and IRIT/CNRS in Toulouse.

## 7. Partnerships and Cooperations

#### 7.1. Regional Initiatives

- Ayman Aljarbouh's PhD (see Section 5.1.3) was partially funded by an ARED grant of the Brittany Regional Council. His doctoral work took place in the context of the Modrio (completed in 2016) and Sys2Soft (completed in 2015) projects on hybrid systems modeling. Ayman Aljarbouh is working on accelerated simulation techniques for hybrid systems. In particular, he is focusing on the regularisation, at runtime, of chattering behaviour and the approximation of Zeno behaviour.
- Benoît Caillaud and Aurélien Lamercerie are participating to the S3PM and SUNSET projects of the CominLabs excellence laboratory <sup>0</sup>. This project focuses on the computation of surgical procedural knowledge models from recordings of individual procedures, and their execution [28]. The objective is to develop an enabling technology for procedural knowledge based computer assistance of surgery. In this project, we demonstrate its potential added value in nurse and surgeon training [36], [35]. In 2017, Benoît Caillaud and Aurélien Lamercerie have released Demodocos, a software synthesizing surgical process models from instances of surgical procedures.

<sup>&</sup>lt;sup>0</sup>http://www.s3pm.cominlabs.ueb.eu/

#### 7.2. National Initiatives

# 7.2.1. Inria Project Lab (IPL): ModeliScale, Languages and Compilation for Cyber-Physical System Design

The project gathers researchers from three Inria teams, and from three other research labs in Grenoble and Paris area.

Name	Team	Inria Center or Laboratory
Vincent Acary	Bipop	Inria Grenoble Rhône Alpes
Albert Benveniste	Hycomes	Inria Rennes
Benoît Caillaud		Bretagne Atlantique
Khalil Ghorbal		
Marc Pouzet	Parkas	ENS
Tim Bourke		Inria Paris
Goran Frehse	Tempo	Verimag-univ. Grenoble Alpes
Antoine Girard		L2S-CNRS, Saclay
Eric Goubault	Cosynus	LIX, École Polytechnique,
Sylvie Putot		Saclay

The main objective of ModeliScale is to advance modeling technologies (languages, compile-time analyses, simulation techniques) for CPS combining physical interactions, communication layers and software components. We believe that mastering CPS comprising thousands to millions of components requires radical changes of paradigms. For instance, modeling techniques must be revised, especially when physics is involved. Modeling languages must be enhanced to cope with larger models. This can only be done by combining new compilation techniques (to master the structural complexity of models) with new mathematical tools (new numerical methods, in particular).

MiodeliScale gathers a broad scope of experts in programming language design and compilation (reactive synchronous programming), numerical solvers (nonsmooth dynamical systems) and hybrid systems modeling and analysis (guaranteed simulation, verification). The research program is carried out in close cooperation with the Modelica community as well as industrial partners, namely, Dassault Systèmes as a Modelica/FMI tool vendor, and EDF and Engie as end users.

## 8. Dissemination

#### 8.1. Promoting Scientific Activities

#### 8.1.1. Scientific Events Organisation

8.1.1.1. General Chair, Scientific Chair

Benoît Caillaud has organized the Synchron'17 open workshop on *Synchronous Programming Languages*<sup>0</sup> that took place at Inria Rennes from November 27th-30th 2017.

#### 8.1.2. Scientific Events Selection

8.1.2.1. Member of the Conference Program Committees

Khalil Ghorbal served as a PC member in the Repeatability Evaluation Committee of HSCC (Hybrid Systems: Computation and Control) 2017.

Albert Benveniste served as a PC member of the International Modelica Conference 2017.

Benoît Caillaud has served on the Steering and Programme Committees of the ACSD'17 conference.

<sup>&</sup>lt;sup>0</sup>https://synchron17.inria.fr

#### 8.1.2.2. Reviewer

Khalil Ghorbal reviewed a paper for the IEEE Conference on Decision and Control 2017.

Albert Benveniste reviewed a paper for FoSSaCS (International Conference on Foundations of Software Science and Computation Structures) 2017.

Benoît Caillaud has reviewed one paper for the LICS'17 conference.

#### 8.1.3. Journal

#### 8.1.3.1. Reviewer - Reviewing Activities

Khalil Ghorbal reviewed a journal paper for the IEEE Transactions on Automatic Control.

Albert Benveniste reviewed a journal paper for the Science of Computer Programming journal.

Benoît Caillaud has reviewed papers for th IEEE Transactions on Control Systems Technology.

#### 8.1.4. Invited Talks

Khalil Ghorbal was invited by Saman Zonouz. Rutgers University, NJ, USA.

Albert Benveniste gave an invited talk at the Laboratory for Information & Decision Systems, MIT, Cambridge, MA, USA.

#### 8.1.5. Scientific Expertise

Albert Benveniste was a reviewer for the ERC Advanced Grant proposals 2017.

#### 8.1.6. Research Administration

Benoît Caillaud is head of the *Language and Software Engineering Department* of IRISA (UMR 6074). The department is composed of 9 research teams and about 120 researchers and students, in Brest, Rennes and Vannes.

#### 8.2. Teaching - Supervision - Juries

#### 8.2.1. Teaching

Master : Benoît Caillaud is teaching with Marc Pouzet a first year master degree course on *hybrid systems modeling*. The course is open to the students registered to the computer science research and innovation curriculum of the university of Rennes 1 and ENS Rennes, France.

Master : Khalil Ghorbal, *Analyse et Conception Formelles*, M1, (chargé de TD), 22h EqTD, University Rennes 1 and ENS Rennes, France

Master : Khalil Ghorbal, Solvers Principle and Architectures, M2, (enseignant principal), 30h EqTD, ENS Rennes, France

Master : Khalil Ghorbal, Modeling Physics with Differential-Algebraic Equations, M2, (enseignant principal), 25h EqTD, Ecole Polytechnique, Palaiseau, France

#### 8.2.2. Supervision

PhD : Ayman Aljarbouh, Accelerated Simulation of Hybrid Systems: Method combining static analysis and runtime execution analysis, University of Rennes 1, defended 13/09/2017, supervised by Benoît Caillaud.

PhD : Guillaume Baudart, *A Synchronous Approach to Quasi-Periodic Systems*, Ecole Normale Superieure (Paris), defended 13/03/2017, co-supervised by Albert Benveniste.

#### 8.2.3. Juries

Benoît Caillaud has been president of PhD defence jury of Mohamed Amine Aouadhi, on 29 September 2017, at LS2N, the University of Nantes, France.

Albert Benveniste participated in the jury of the PhD thesis of Guillaume Baudart.

## 9. Bibliography

#### Major publications by the team in recent years

- [1] A. BENVENISTE, T. BOURKE, B. CAILLAUD, M. POUZET.*Non-standard semantics of hybrid systems modelers*, in "Journal of Computer and System Sciences", 2012, vol. 78, n<sup>o</sup> 3, p. 877-910, This work was supported by the SYNCHRONICS large scale initiative of Inria [DOI: 10.1016/J.JCSS.2011.08.009], http://hal.inria.fr/hal-00766726.
- [2] A. BENVENISTE, B. CAILLAUD, D. NICKOVIC, R. PASSERONE, J.-B. RACLET, P. REINKEMEIER, A. L. SANGIOVANNI-VINCENTELLI, W. DAMM, T. A. HENZINGER, K. G. LARSEN. Contracts for System Design, Inria, November 2012, n<sup>o</sup> RR-8147, 65, http://hal.inria.fr/hal-00757488.
- [3] J.-B. RACLET, E. BADOUEL, A. BENVENISTE, B. CAILLAUD, A. LEGAY, R. PASSERONE. A Modal Interface Theory for Component-based Design, in "Fundamenta Informaticae", 2011, vol. 108, n<sup>o</sup> 1-2, p. 119-149, http://dx.doi.org/10.3233/FI-2011-416.

#### **Publications of the year**

#### **Doctoral Dissertations and Habilitation Theses**

[4] A. ALJARBOUH. Accelerated Simulation of Hybrid Systems : Method combining static analysis and run-time execution analysis, Université de Rennes 1, France, September 2017, https://hal.inria.fr/tel-01614081.

#### **Articles in International Peer-Reviewed Journal**

- [5] A. BENVENISTE, T. BOURKE, B. CAILLAUD, B. PAGANO, M. POUZET. A Type-based Analysis of Causality Loops in Hybrid Systems Modelers, in "Nonlinear Analysis: Hybrid Systems", November 2017, vol. 26, p. 168–189 [DOI: 10.1016/J.NAHS.2017.04.004], https://hal.inria.fr/hal-01549183.
- [6] S. MITSCH, K. GHORBAL, D. VOGELBACHER, A. PLATZER. Formal verification of obstacle avoidance and navigation of ground robots, in "International Journal of Robotics Research", 2017, vol. 36, n<sup>o</sup> 12, p. 1312–1340 [DOI: 10.1177/0278364917733549], https://hal.inria.fr/hal-01658197.
- [7] A. SOGOKON, K. GHORBAL, T. T. JOHNSON. Operational Models for Piecewise-Smooth Systems, in "ACM Transactions on Embedded Computing Systems (TECS)", October 2017, vol. 16, n<sup>o</sup> 5s, p. 185:1–185:19 [DOI: 10.1145/3126506], https://hal.inria.fr/hal-01658196.

#### **International Conferences with Proceedings**

[8] A. BENVENISTE, B. CAILLAUD, H. ELMQVIST, K. GHORBAL, M. OTTER, M. POUZET. Structural Analysis of Multi-Mode DAE Systems, in "Proceedings of the 20th International Conference on Hybrid Systems: Computation and Control, HSCC 2017", Pittsburgh, PA, United States, April 2017 [DOI: 10.1145/3049797.3049806], https://hal.inria.fr/hal-01521918. [9] B. MARTIN, K. GHORBAL, E. GOUBAULT, S. PUTOT. Formal Verification of Station Keeping Maneuvers for a Planar Autonomous Hybrid System, in "FVAV 2017 - 1st Formal Verification of Autonomous Vehicles Workshop", Turin, Italy, L. BULWAHN, M. KAMALI, S. LINKER (editors), FVAV@iFM 2017, September 2017, vol. 257, p. 91–104 [DOI: 10.4204/EPTCS.257.9], https://hal.archives-ouvertes.fr/hal-01657848.

#### Scientific Books (or Scientific Book chapters)

[10] A. BENVENISTE, B. CAILLAUD.Synchronous Interfaces and Assume/Guarantee Contracts, in "Models, Algorithms, Logics and Tools - Essays Dedicated to Kim Guldstrand Larsen on the Occasion of His 60th Birthday", L. ACETO, G. BACCI, G. BACCI, A. INGÓLFSDÓTTIR, R. MARDARE (editors), Theoretical Computer Science and General Issues, Springer, July 2017, vol. 10460, p. 233-248 [DOI : 10.1007/978-3-319-63121-9\_12], https://hal.inria.fr/hal-01616369.

#### **Research Reports**

[11] A. BENVENISTE, B. CAILLAUD, H. ELMQVIST, K. GHORBAL, M. OTTER, M. POUZET. *Structural Analysis of Multi-Mode DAE Systems*, Inria, 2017, n<sup>o</sup> RR-8933, p. 1-23, https://hal.inria.fr/hal-01343967.

#### **References in notes**

- [12] N. J. CUTLAND (editor). Nonstandard analysis and its applications, Cambridge Univ. Press, 1988.
- [13] IEEE Standard VHDL Analog and Mixed-Signal Extensions, Std 1076.1-1999, 1999, http://dx.doi.org/10. 1109/IEEESTD.1999.90578.
- [14] A. ANTONIK, M. HUTH, K. G. LARSEN, U. NYMAN, A. WASOWSKI.20 Years of Modal and Mixed Specifications, in "Bulletin of European Association of Theoretical Computer Science", 2008, vol. 1, n<sup>o</sup> 94.
- [15] C. BAIER, J.-P. KATOEN. Principles of Model Checking, MIT Press, Cambridge, 2008.
- [16] A. BENVENISTE, T. BOURKE, B. CAILLAUD, B. PAGANO, M. POUZET. A Type-Based Analysis of Causality Loops In Hybrid Systems Modelers, December 2013, Deliverable D3.1\_1 v 1.0 of the Sys2soft collaborative project "Physics Aware Software", https://hal.inria.fr/hal-00938866.
- [17] A. BENVENISTE, T. BOURKE, B. CAILLAUD, M. POUZET. *Semantics of multi-mode DAE systems*, August 2013, Deliverable D.4.1.1 of the ITEA2 Modrio collaborative project, https://hal.inria.fr/hal-00938891.
- [18] A. BENVENISTE, B. CAILLAUD, H. ELMQVIST, K. GHORBAL, M. OTTER, M. POUZET. *Structural Analysis of Multi-Mode DAE Systems*, in "Proc. of the 20th ACM International Conference on Hybrid Systems: Computation and Control, HSCC'17", Pittsburgh, PA, USA, April 2017, to appear.
- [19] A. BENVENISTE, B. CAILLAUD, A. FERRARI, L. MANGERUCA, R. PASSERONE, C. SOFRONIS. *Multiple Viewpoint Contract-Based Specification and Design*, in "Proceedings of the Software Technology Concertation on Formal Methods for Components and Objects (FMCO'07)", Amsterdam, The Netherlands, Revised Lectures, Lecture Notes in Computer Science, Springer, October 2008, vol. 5382.
- [20] A. BENVENISTE, B. CAILLAUD, B. PAGANO, M. POUZET. A type-based analysis of causality loops in hybrid modelers, in "HSCC '14: International Conference on Hybrid Systems: Computation and Control", Berlin, Germany, Proceedings of the 17th international conference on Hybrid systems: computation and

control (HSCC '14), ACM Press, April 2014, 13 [DOI : 10.1145/2562059.2562125], https://hal.inria.fr/hal-01093388.

- [21] N. BERTRAND, A. LEGAY, S. PINCHINAT, J.-B. RACLET. A Compositional Approach on Modal Specifications for Timed Systems, in "11th International Conference on Formal Engineering Methods (ICFEM'09)", Rio de Janeiro, Brazil, LNCS, Springer, December 2009, vol. 5885, p. 679-697, http://hal.inria.fr/inria-00424356/ en.
- [22] N. BERTRAND, A. LEGAY, S. PINCHINAT, J.-B. RACLET.*Modal event-clock specifications for timed component-based design*, in "Science of Computer Programming", 2011, http://dx.doi.org/10.1016/j.scico. 2011.01.007.
- [23] N. BERTRAND, S. PINCHINAT, J.-B. RACLET. Refinement and Consistency of Timed Modal Specifications, in "3rd International Conference on Language and Automata Theory and Applications (LATA'09)", Tarragona, Spain, LNCS, Springer, April 2009, vol. 5457, p. 152-163 [DOI: 10.1007/978-3-642-00982-2\_13], http:// hal.inria.fr/inria-00424283/en.
- [24] P. BHADURI, I. STIERAND.A proposal for real-time interfaces in SPEEDS, in "Design, Automation and Test in Europe (DATE'10)", IEEE, 2010, p. 441-446.
- [25] S. BLIUDZE. Un cadre formel pour l'étude des systèmes industriels complexes: un exemple basé sur l'infrastructure de l'UMTS, Ecole Polytechnique, 2006.
- [26] S. BLIUDZE, D. KROB.Modelling of Complex Systems: Systems as Dataflow Machines, in "Fundam. Inform.", 2009, vol. 91, n<sup>o</sup> 2, p. 251–274.
- [27] G. BOUDOL, K. G. LARSEN. Graphical Versus Logical Specifications, in "Theor. Comput. Sci.", 1992, vol. 106, n<sup>o</sup> 1, p. 3-20.
- [28] B. CAILLAUD.Surgical Process Mining with Test and Flip Net Synthesis, in "Application of Region Theory (ART)", Barcelona, Spain, R. BERGENTHUM, J. CARMONA (editors), July 2013, p. 43-54, http://hal.inria.fr/ hal-00872284.
- [29] B. CAILLAUD, B. DELAHAYE, K. G. LARSEN, A. LEGAY, M. L. PEDERSEN, A. WASOWSKI. Compositional design methodology with constraint Markov chains, in "QEST 2010", Williamsburg, Virginia, United States, September 2010 [DOI : 10.1109/QEST.2010.23], http://hal.inria.fr/inria-00591578/en.
- [30] B. CAILLAUD, B. DELAHAYE, K. G. LARSEN, A. LEGAY, M. L. PEDERSEN, A. WASOWSKI. Constraint Markov Chains, in "Theoretical Computer Science", May 2011, vol. 412, n<sup>o</sup> 34, p. 4373-4404 [DOI: 10.1016/J.TCS.2011.05.010], http://hal.inria.fr/hal-00654003/en.
- [31] A. CHAKRABARTI.A Framework for Compositional Design and Analysis of Systems, EECS Department, University of California, Berkeley, Dec 2007, http://www.eecs.berkeley.edu/Pubs/TechRpts/2007/EECS-2007-174.html.
- [32] A. CHAKRABARTI, L. DE ALFARO, T. A. HENZINGER, M. STOELINGA. Resource Interfaces, in "EMSOFT", R. ALUR, I. LEE (editors), Lecture Notes in Computer Science, Springer, 2003, vol. 2855, p. 117-133.

- [33] E. Y. CHANG, Z. MANNA, A. PNUELI. *Characterization of Temporal Property Classes*, in "ICALP", W. KUICH (editor), Lecture Notes in Computer Science, Springer, 1992, vol. 623, p. 474-486.
- [34] E. CLARKE, O. GRUMBERG, D. PELED. Model Checking, MIT Press, 1999.
- [35] G. CLAUDE, V. GOURANTON, B. CAILLAUD, B. GIBAUD, B. ARNALDI, P. JANNIN. Synthesis and Simulation of Surgical Process Models, in "Studies in Health Technology and Informatics", 2016, vol. 220, p. 63–70 [DOI: 10.3233/978-1-61499-625-5-63], https://hal.archives-ouvertes.fr/hal-01300990.
- [36] G. CLAUDE, V. GOURANTON, B. CAILLAUD, B. GIBAUD, P. JANNIN, B. ARNALDI. From Observations to Collaborative Simulation: Application to Surgical Training, in "ICAT-EGVE 2016 - International Conference on Artificial Reality and Telexistence, Eurographics Symposium on Virtual Environments", Little Rock, Arkansas, United States, December 2016, https://hal.archives-ouvertes.fr/hal-01391776.
- [37] W. DAMM, E. THADEN, I. STIERAND, T. PEIKENKAMP, H. HUNGAR. Using Contract-Based Component Specifications for Virtual Integration and Architecture Design, in "Proceedings of the 2011 Design, Automation and Test in Europe (DATE'11)", March 2011.
- [38] A. DAVID, K. G. LARSEN, A. LEGAY, U. NYMAN, A. WASOWSKI.ECDAR: An Environment for Compositional Design and Analysis of Real Time Systems, in "Automated Technology for Verification and Analysis -8th International Symposium, ATVA 2010, Singapore, September 21-24, 2010. Proceedings", 2010, p. 365-370.
- [39] A. DAVID, K. G. LARSEN, A. LEGAY, U. NYMAN, A. WASOWSKI.*Timed I/O automata: a complete specification theory for real-time systems*, in "Proceedings of the 13th ACM International Conference on Hybrid Systems: Computation and Control, HSCC 2010, Stockholm, Sweden, April 12-15, 2010", 2010, p. 91-100.
- [40] B. DELAHAYE, J.-P. KATOEN, K. G. LARSEN, A. LEGAY, M. L. PEDERSEN, F. SHER, A. WA-SOWSKI. Abstract Probabilistic Automata, in "VMCAI", R. JHALA, D. A. SCHMIDT (editors), Lecture Notes in Computer Science, Springer, 2011, vol. 6538, p. 324-339.
- [41] F. DIENER, G. REEB. Analyse non standard, Hermann, 1989.
- [42] D. L. DILL.Trace Theory for Automatic Hierarchical Verification of Speed-Independent Circuits, ACM Distinguished Dissertations, MIT Press, 1989.
- [43] Y. IWASAKI, A. FARQUHAR, V. SARASWAT, D. BOBROW, V. GUPTA. Modeling Time in Hybrid Systems: How Fast Is "Instantaneous"?, in "IJCAI", 1995, p. 1773–1781.
- [44] L. LAMPORT. *Proving the Correctness of Multiprocess Programs*, in "IEEE Trans. Software Eng.", 1977, vol. 3, n<sup>o</sup> 2, p. 125-143.
- [45] K. G. LARSEN, U. NYMAN, A. WASOWSKI. On Modal Refinement and Consistency, in "Proc. of the 18th International Conference on Concurrency Theory (CONCUR'07)", Springer, 2007, p. 105–119.
- [46] K. G. LARSEN, B. THOMSEN. *A Modal Process Logic*, in "Proceedings of the Third Annual Symposium on Logic in Computer Science (LICS'88)", IEEE, 1988, p. 203-210.

- [47] T. LINDSTRØM. An Invitation to Nonstandard Analysis, in "Nonstandard Analysis and its Applications", N. J. CUTLAND (editor), Cambridge Univ. Press, 1988, p. 1–105.
- [48] N. A. LYNCH. Input/Output Automata: Basic, Timed, Hybrid, Probabilistic, Dynamic, ..., in "CONCUR", R. M. AMADIO, D. LUGIEZ (editors), Lecture Notes in Computer Science, Springer, 2003, vol. 2761, p. 187-188.
- [49] N. A. LYNCH, E. W. STARK. A Proof of the Kahn Principle for Input/Output Automata, in "Inf. Comput.", 1989, vol. 82, n<sup>o</sup> 1, p. 81-92.
- [50] Z. MANNA, A. PNUELI. Temporal verification of reactive systems: Safety, Springer, 1995.
- [51] B. MEYER. Applying "Design by Contract", in "Computer", October 1992, vol. 25, n<sup>o</sup> 10, p. 40–51, http://dx. doi.org/10.1109/2.161279.
- [52] P. NUZZO, A. L. SANGIOVANNI-VINCENTELLI, X. SUN, A. PUGGELLI. Methodology for the Design of Analog Integrated Interfaces Using Contracts, in "IEEE Sensors Journal", Dec. 2012, vol. 12, n<sup>o</sup> 12, p. 3329–3345.
- [53] A. ROBINSON. Non-Standard Analysis, Princeton Landmarks in Mathematics, 1996, ISBN 0-691-04490-2.
- [54] E. SIKORA, B. TENBERGEN, K. POHL. Industry needs and research directions in requirements engineering for embedded systems, in "Requirements Engineering", 2012, vol. 17, p. 57–78, http://link.springer.com/ article/10.1007/s00766-011-0144-x.
- [55] L. DE ALFARO. Game Models for Open Systems, in "Verification: Theory and Practice", Lecture Notes in Computer Science, Springer, 2003, vol. 2772, p. 269-289.
- [56] L. DE ALFARO, T. A. HENZINGER. *Interface automata*, in "Proc. of the 9th ACM SIGSOFT International Symposium on Foundations of Software Engineering (FSE'01)", ACM Press, 2001, p. 109–120.
- [57] L. DE ALFARO, T. A. HENZINGER. Interface-based design, in "In Engineering Theories of Software Intensive Systems, proceedings of the Marktoberdorf Summer School", Kluwer, 2004.
- [58] L. DE ALFARO, T. A. HENZINGER, M. STOELINGA.*Timed Interfaces*, in "Proc. of the 2nd International Workshop on Embedded Software (EMSOFT'02)", Lecture Notes in Computer Science, Springer, 2002, vol. 2491, p. 108–122.

# **Project-Team I4S**

# Statistical Inference for Structural Health Monitoring

IN PARTNERSHIP WITH: Institut français des sciences et technologies des transports, de l'aménagement et des réseaux - IFSTTAR

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Optimization and control of dynamic systems

## **Table of contents**

1.	Personnel	. 431	
2.	2. Overall Objectives		
	2.1.1. In Summary		
	2.1.2. Introduction to physics driven dynamical models in the context of civil engineering ela		
	structures	433	
	2.1.2.1. Multi-fold thermal effects	434	
	2.1.2.2. Toward a multidisciplinary approach	434	
	2.1.2.3. Models for monitoring under environmental changes - scientific background	435	
3.	Research Program	. 437	
	3.1. Vibration analysis	437	
	3.1.1. Identification	438	
	3.1.2. Detection	439	
	3.1.3. Diagnostics	440	
	3.2. Thermal methods	441	
	3.2.1. Infrared thermography and heat transfer	441	
	3.2.1.1. Infrared radiation	441	
	3.2.1.2. Infrared Thermography	443	
	3.2.2. Heat transfer theory	443	
	3.2.3. Inverse model for parameters estimation	444	
	3.3. Reflectometry-based methods for electrical engineering and for civil engineering	445	
	3.3.1. Mathematical model of electric cables and networks	445	
	3.3.2. The inverse scattering theory applied to cables	446	
	3.4. Research Program	447	
	3.4.1. Vibration analysis and monitoring	448	
	3.4.1.1. Direct vibration modeling under temperature changes	448	
	3.4.1.2. Damage localization algorithms (in the case of localized damages such as crack	s)448	
	3.4.1.3. Uncertainty quantification for system identification algorithms	448	
	3.4.2. Reflectometry-based methods for civil engineering structure health monitoring	448	
	3.4.3. Non Destructive testing of CFRP bonded on concrete through active thermography	449	
	3.4.4. IRSHM: Multi-Sensing system for outdoor thermal monitoring	449	
	3.4.5. R5G: The 5th Generation Road	449	
4.	Application Domains	. 450	
	4.1. Civil Engineering	450	
_	4.2. Electrical cable and network monitoring	451	
5.	Highlights of the Year	. 451	
6.	New Software and Platforms	. 451	
	6.1. PEGASE	451	
_	6.2. TrackingMecaSysEvo	452	
7.		. 453	
	7.1. Outdoor InfraRed Thermography	453	
	7.1.1. Joint thermal and electromagnetic diagnostics	453	
	7.1.2. Long term monitoring of transport infrastructures: from deployment to standardization	1 453	
	7.1.3. Infrared data reconstruction and calibration for long term monitoring	453	
	7.2. Data management of Smart territories and cities	453	
	7.3. Smarts roads and R5G	453	
	7.3.1. Energy exchange modelization and infrared monitoring for hybrid pavement structure	453	
	7.3.2. Phase change materials characterization	454	
	7.4. Methods for building performance assessment	454	
	/.5. System identification	454	

7.5.1.	Variance estimation of modal parameters from subspace-based system identification	454
7.5.2.	Bayesian parameter estimation for parameter varying systems using interacting Kalı	nan
	filters	454
7.5.3.	From structurally independent local LTI models to LPV model	455
7.5.4.	Stability of the Kalman filter for output error systems	455
7.5.5.	Reduced-order interval-observer design for dynamic systems with time-invariant un	cer-
	tainty	455
7.5.6.	Parameter uncertainties quantification for finite element based subspace fitting approact	ches
		455
7.6. Da	amage diagnosis	456
7.6.1.	Damage detection by perturbation analysis and additive change detection theory	456
7.6.2.	Damage localization using the statistical subspace damage localization method	456
7.6.3.	Stochastic Subspace-Based Damage Detection with Uncertainty in the Reference N	Null
	Space	456
7.6.4.	Statistical damage localization with stochastic load vectors	457
7.6.5.	Transfer matrices-based statistical damage localization and quantification	457
7.6.6.	Statistical damage localization based on Mahalanobis distance	457
7.6.7.	On the value of Information for SHM	457
7.6.8.	Structural system reliability and damage detection information	458
7.6.9.	Estimation of a cable resistance profile with readaptation of mismatched measuren	nent
	instrument	458
7.7. Se	nsor and hardware based research	458
7.7.1.	Cracks detection in pavement by a distributed fiber optic sensing technology	458
7.7.2.	Wireless sensors and GPS synchronization	459
Bilateral	Contracts and Grants with Industry	. 459
8.1. Bi	lateral Contracts with Industry	459
8.1.1.	Contracts with SVS	459
8.1.2.	Contract with SNCF: DEMETER	459
8.1.3.	Contracts with SDEL-CC (VINCI Group)	460
8.2. Bi	lateral Grants with Industry	461
Partners	hips and Cooperations	. 461
9.1. Re	egional Initiatives	461
9.1.1.	MONEOL	461
9.1.2.	Interactive Communication (InterCom): Massive random access to subsets of compres	ssed
	correlated data	461
9.1.3.	MAG2C-Pont Tabarly	462
9.1.4.	MAG2C-MOSIWIND (MOnitoring of Structural Integrity of an onshore WIND turk	oine
	slab foundation and tower)	462
9.1.5.	Collaboration with GeM	463
9.1.6.	Collaboration with IETR	463
9.2. Na	ational Initiatives	463
9.2.1.	High speed rail track instrumentation	463
9.2.2.	ANR Resbati	463
9.2.3.	Equipex Sense-City	463
9.3. Eu	ropean Initiatives	464
9.3.1.	FP7 & H2020 Projects	464
9.3	3.1.1. Built to Specifications (Built2Spec)	464
9.3	3.1.2. INFRASTAR(Innovation and Networking for Fatigue and Reliability Analysi	s of
	Structures – Training for Assessment of Risk)	464
9.3.2.	Collaborations in European Programs, Except FP7 & H2020	465
9.3	3.2.1. COST Action TU 1402	465

8.

9.

	9.	3.2.2.	PROCOPE 37826QE	465
	9.3.3.	Colla	borations with Major European Organizations	465
	9.3.4.	Other	r European Programs	465
	9.4. In	ternatio	nal Initiatives	466
	9.4	4.1.1.	Collaboration with CNR, Italy	466
	9.4	4.1.2.	Collaboration with British Columbia University, Canada	466
	9.4	4.1.3.	Collaboration with BAM, Germany	466
	9.4	4.1.4.	Collaboration with Politecnico di Milano, Italy	466
	9.4	4.1.5.	Collaboration with Technical University of Denmark (DTU)	466
	9.4	4.1.6.	Collaboration with Aalborg University, Denmark	466
	9.5. In	ternation	nal Research Visitors	466
	9.5.1.	Visits	s of International Scientists	466
	9.5.2.	Visits	s to International Teams	466
10.	Dissem	ination		
	10.1. Pr	omoting	g Scientific Activities	467
	10.1.1	. Scien	tific Events Selection	467
	10	).1.1.1.	Member of the Conference Program Committees	467
	10	).1.1.2.	Reviewer	467
	10.1.2. Journal		467	
	10	.1.2.1.	Member of the Editorial Boards	467
	10	0.1.2.2.	Reviewer - Reviewing Activities	468
	10.1.3	. Invite	ed Talks	468
	10.1.4. Research Administration			468
	10.2. Te	eaching	- Supervision - Juries	468
	10.2.1	. Teacl	hing	468
	10.2.2. Supervision			469
	10.2.3	Juries	S	470
	10.3. Po	opulariza	ation	470
11.	Bibliog	raphy		470

#### **Project-Team I4S**

*Creation of the Team: 2009 April 01, updated into Project-Team: 2013 January 01* **Keywords:** 

#### **Computer Science and Digital Science:**

A6.1.5. - Multiphysics modeling

A6.2.1. - Numerical analysis of PDE and ODE

- A6.2.4. Statistical methods
- A6.2.5. Numerical Linear Algebra
- A6.2.6. Optimization
- A6.3.1. Inverse problems
- A6.3.3. Data processing
- A6.3.4. Model reduction
- A6.3.5. Uncertainty Quantification
- A6.4.3. Observability and Controlability

#### **Other Research Topics and Application Domains:**

- B3.1. Sustainable development
- B3.2. Climate and meteorology
- B3.3.1. Earth and subsoil
- B4.3.2. Hydro-energy
- B4.3.3. Wind energy
- B4.3.4. Solar Energy
- B5.1. Factory of the future
- B5.2. Design and manufacturing
- B5.9. Industrial maintenance
- B6.5. Information systems
- B7.2.2. Smart road
- B8.1. Smart building/home
- B8.1.1. Energy for smart buildings
- B8.1.2. Sensor networks for smart buildings
- B8.2. Connected city

## 1. Personnel

#### **Research Scientists**

Laurent Mevel [Team leader, Inria, Senior Researcher, HDR] Jean Dumoulin [IFSTTAR, Researcher, scientific co-head of I4S] Qinghua Zhang [Inria, Senior Researcher, HDR] Xavier Chapeleau [IFSTTAR, Researcher] Michael Doehler [Inria, Researcher]

#### **Faculty Member**

Frederic Gillot [Ecole centrale de Lyon, Associate Professor, since September 2017]

#### **Technical Staff**

Antoine Crinière [Inria] Quentin Bossard [IFSTTAR] Ivan Guéguen [IFSTTAR] Vincent Le Cam [IFSTTAR, Researcher, Team leader of SII Lab at IFSTTAR] Mathieu Le Pen [IFSTTAR]

#### **PhD Students**

Saeid Allahdadian [UBC Vancouver, until June 2017] Nassif Berrabah [EDF, until November 2017] Md Delwar Hossain Bhuyan [Inria, until November 2017] Nicolas Le Touz [IFSTTAR] David Pallier [IFSTTAR, since November 2017] Antoine Bassil [IFSTTAR] Thibaud Toullier [Inria] Eva Viefhues [BAM] Francesco Giordano [Politecnico di Milano, since November 2017]

#### **Post-Doctoral Fellows**

Guillaume Gautier [Inria since September 2017, CEA-tech until August 2017] Ludovic Gaverina [IFSTTAR] Subhamoy Sen [Inria, until May 2017]

#### Administrative Assistants

Helene de La Ruée [Inria] Yveline Gourbil [IFSTTAR]

## 2. Overall Objectives

#### 2.1. Overall Objectives

monitoring, system identification, on-line identification and detection algorithms, statistical hypotheses testing, reflectometry, infrared thermography, non destructive testing, sensors fusion, optimal sensors placement, vibration-based structural analysis and damage detection and localization, aeronautics, civil engineering

#### 2.1.1. In Summary

The objective of this team is the development of Structural Health Monitoring techniques by intrinsic coupling of statistics and thermo-aeroelastic mixing modeling for the development of robust and autonomous structural health monitoring solutions of mechanical structures. The emphasis of the team is the handling of very large systems such as the recent wind energy converters currently being installed in Europe, building on the expertise acquired by the team on bridges as an example of civil engineering structure, and for aircrafts and helicopters in the context of aero elastic instability monitoring. The necessity of system identification and damage detection systems robust to environmental variations and being designed to handle a very large model dimension motivates us. As examples, the explosion in the installed number of sensors and the robustness to temperature variation will be the main focus of the team. This implies new statistical and numerical technologies as well as improvements on the modeling of the underlying physical models. Many techniques and methods originate from the mechanical community and thus exhibit a very deep understanding of the underlying physics and mechanical behavior of the structure. On the other side, system identification techniques developed within the control community are more related to data modeling and take into account the underlying random nature of measurement noise. Bringing these two communities together is the objective of this joint team between Inria and IFSTTAR. It will results hopefully in methods numerically robust, statistically efficient and also mixing modeling of both the uncertainties related to the data and the associated complex physical models related to the laws of physics and finite element models.
Damage detection in civil structures has been a main focus over the last decade. Still, those techniques need to be matured to be operable and installed on structures in operation, and thus be robust to environmental nuisances. Then, damage localization, quantification and prognosis should be in that order addressed by the team. To be precise and efficient, it requires correct mixing between signal processing, statistical analysis, Finite Elements Models (FEM) updating and a yet to be available precise modeling of the environmental effects such as temperature through 3D field reconstruction.

Theoretical and practical questions are more and more complex. For example, in civil engineering, from handling hundreds of sensors automatically during some long period of time to localize and quantify damage with or without numerical models. Very large heavily instrumented structures are yet to come and they will ask for a paradigm in how we treat them from a renewed point of view. As the structures become large and complex, also the thermal and aeroelastic (among others) models become complex. Bridges and aircrafts are the main focus of our research. Opening our expertise on new applications topics such as helicopters and wind energy converters is also part of our priorities.

### 2.1.1.1. Objectives

The main objectives of the team are first to pursue current algorithmic research activities, in order to accommodate still-to-be-developed complex physical models. More precisely, we want successively

- To develop statistical algorithms robust to noise and variation in the environment
- To handle transient and highly varying systems under operational conditions
- To consider the impact of uncertainties on the current available identification algorithms and develop efficient, robust and fast implementation of such quantities
- To consider relevant non trivial thermal models for usage in rejection based structural health monitoring and more generally to mix numerical model, physical modeling and data
- To develop theoretical and software tools for monitoring and localization of damages on civil structures or instability for aircrafts
- To explore new paradigms for handling very large and complex structures heavily instrumented (distributed computing)
- To study the characteristics of the monitored mechanic structures in terms of electromagnetic propagation, in order to develop monitoring methods based on electrical instrumentations.
- To consider society concerns (damage quantification and remaining life prognosis)

# 2.1.2. Introduction to physics driven dynamical models in the context of civil engineering elastic structures

The design and maintenance of flexible structures subject to noise and vibrations is an important topic in civil and mechanical engineering. It is an important component of comfort (cars and buildings) and contributes significantly to the safety related aspects of design and maintenance (aircrafts, aerospace vehicles and payloads, long-span bridges, high-rise towers...). Requirements from these application areas are numerous and demanding.

Detailed physical models derived from first principles are developed as part of system design. These models involve the dynamics of vibrations, sometimes complemented by other physical aspects (fluid-structure interaction, aerodynamics, thermodynamics).

Laboratory and in-operation tests are performed on mock-up or real structures, in order to get so-called modal models, ie to extract the modes and damping factors (these correspond to system poles), the mode shapes (corresponding eigenvectors), and loads. These results are used for updating the design model for a better fit to data, and sometimes for certification purposes (e.g. in flight domain opening for new aircrafts, reception for large bridges).

The monitoring of structures is an important activity for the system maintenance and health assessment. This is particularly important for civil structures. Damaged structures would typically exhibit often very small changes in their stiffness due to the occurrence of cracks, loss of prestressing or post tensioning, chemical reactions, evolution of the bearing behavior and most importantly scour. A key difficulty is that such system characteristics are also sensitive to environmental conditions, such as temperature effects (for civil structures), or external loads (for aircrafts). In fact these environmental effects usually dominate the effect of damage. This is why, for very critical structures such as aircrafts, detailed active inspection of the structures is performed as part of the maintenance. Of course, whenever modal information is used to localize a damage, the localization of a damage should be expressed in terms of the physical model, not in terms of the modal model used in system identification. Consequently, the following elements are encountered and must be jointly dealt with when addressing these applications: design models from the system physics, modal models used in structural identification, and, of course, data from sensors. Corresponding characteristics are given now: Design models are Finite Element models, sometimes with tens or hundreds of thousands elements, depending on professional habits which may vary from one sector to another. These models are linear if only small vibrations are considered; still, these models can be large if medium-frequency spectrum of the load is significant. In addition, nonlinearities enter as soon as large vibrations or other physical effects (aerodynamics, thermodynamics, ...) are considered. Moreover stress-strain paths and therefore the response (and load) history comes into play.

Sensors can range from a handful of accelerometers or strain gauges, to thousands of them, if NEMS ( Nano Electro Mechanical Structures), MEMS (Microelectromechanical systems) or optical fiber sensors are used. Moreover, the sensor output can be a two-dimensional matrix if electro magnet (IR (infrared), SAR, shearography ...) or other imaging technologies are used.

#### 2.1.2.1. Multi-fold thermal effects

The temperature constitutes an often dominant load because it can generate a deflection as important as that due to the self-weight of a bridge. In addition, it sometimes provokes abrupt slips of bridge spans on their bearing devices, which can generate significant transient stresses as well as a permanent deformation, thus contributing to fatigue.

But it is also well-known that the dynamic behavior of structures under monitoring can vary under the influence of several factors, including the temperature variations, because they modify the stiffness and thus the modes of vibration. As a matter of fact, depending on the boundary conditions of the structure, possibly uniform thermal variations can cause very important variations of the spectrum of the structure, up to 10%, because in particular of additional prestressing, not forgetting pre strain, but also because of the temperature dependence of the characteristics of materials. As an example, the stiffness of elastomeric bearing devices vary considerably in the range of extreme temperatures in some countries. Moreover, eigenfrequencies and modal shapes do not depend monotonically with temperature. Abrupt dynamical behavior may show up due to a change of boundary conditions e.g. due to limited expansion or frost bearing devices. The temperature can actually modify the number of contact points between the piles and the main span of the bridge. Thus the environmental effects can be several orders of magnitude more important than the effect of true structural damages. It will be noted that certain direct methods aiming at detecting local curvature variations stumble on the dominating impact of the thermal gradients. In the same way, the robustness and effectiveness of model-based structural control would suffer from any unidentified modification of the vibratory behavior of the structure of interest. Consequently, it is mandatory to cure dynamic sensor outputs from thermal effects before signal processing can help with a diagnostics on the structure itself, otherwise the possibility of reliable ambient vibration monitoring of civil structures remains questionable. Despite the paramount interest this question deserves, thermal elimination still appears to challenge the SHM community.

#### 2.1.2.2. Toward a multidisciplinary approach

Unlike previously mentioned blind approaches, successful endeavours to eliminate the temperature from subspace-based damage detection algorithms prove the relevance of relying on predictive thermo-mechanical models yielding the prestress state and associated strains due to temperature variations. As part of the CONSTRUCTIF project supported by the Action Concertée Incitative Sécurité Informatique of the French

Ministry for Education and Research, very encouraging results in this direction were obtained and published. They were substantiated by laboratory experiments of academic type on a simple beam subjected to a known uniform temperature. Considering the international pressure toward reliable methods for thermal elimination, these preliminary results pave the ground to a new SHM paradigm. Moreover, for one-dimensional problems, it was shown that real time temperature identification based on optimal control theory is possible provided the norm of the reconstructed heat flux is properly chosen. Finally, thermo-mechanical models of vibrating thin structures subject to thermal prestress, prestrain, geometric imperfection and damping have been extensively revisited. This project led by Inria involved IFSTTAR where the experiments were carried out. The project was over in July 2006. Note that thermo-mechanics of bridge piles combined with an *ad hoc* estimation of thermal gradients becomes of interest to practicing engineers. Thus, I4S's approach should suit advanced professional practice. Finite element analysis is also used to predict stresses and displacements of large bridges in Hong-Kong bay .

Temperature rejection is the primary focus and challenge for I4S's SHM projects in civil engineering, like SIMS project in Canada, ISMS in Danemark or SIPRIS in France.

A recent collaboration between Inria and IFSTTAR has demonstrated the efficiency of reflectometry-based methods for health monitoring of some civil engineering structures, notably external post-tensioned cables. Based on a mathematical model of electromagnetic propagation in mechanical structures, the measurement of reflected and transmitted electromagnetic waves by the monitored structures allows to detect structural failures. The interaction of such methods with those based on mechanical and thermal measurements will reinforce the multidisciplinary approach developed in our team.

### 2.1.2.3. Models for monitoring under environmental changes - scientific background

We will be interested in studying linear stochastic systems, more precisely, assume at hand a sequence of observations  $Y_n$  measured during time,

$$\begin{cases} X_{n+1} = AX_n + V_n \\ Y_n = HX_n + W_n \end{cases}$$
(1)

where  $V_n$  and  $W_n$  are zero mean random variables, A is the transition matrix of the system, H is the observation matrix between state and observation, and  $X_n$  the process describing the monitored system.  $X_n$  can be related to a physical process (for example, for a mechanical structure, the collection of displacements and velocities at different points). Different problems arise

1/ identify and characterize the structure of interest. It may be possible by matching a parametric model to the observed time series  $Y_n$  in order to minimize some given criterion, whose minimum will be the best approximation describing the system,

2/ decide if the measured data describe a system in a so called "reference" state (the term "reference" is used in the context of fault detection, where the reference is considered to be safe) and monitor its deviations with respect of its nominal reference state.

Both problems should be addressed differently if

1/ we consider that the allocated time to measurement is large enough, resulting in a sequence of  $Y_n$  whose length tends to infinity, a requirement for obtaining statistical convergence results. It corresponds to the identification and monitoring of a dynamical system with slow variations. For example, this description is well suited to the long-term monitoring of civil structures, where records can be measured during relatively (to sampling rate) large periods of time (typically many minutes or hours).

2/ we are interested in systems, whose dynamic is fast with respect to the sampling rate, most often asking for reaction in terms of seconds. It is, for example, the case for mission critical applications such as in-flight control or real-time security and safety assessment. Both aeronautics and transport or utilities infrastructures are concerned. In this case, fast algorithms with sample-by-sample reaction are necessary.

The monitoring of mechanical structures can not be addressed without taking into account the close environment of the considered system and their interactions. Typically, monitored structures of interest do not reside in laboratory but are considered in operational conditions, undergoing temperature, wind and humidity variations, as well as traffic, water flows and other natural or man-made loads. Those variations do imply a variation of the eigenproperties of the monitored structure, variations to be separated from the damage/instability induced variations.

For example, in civil engineering, an essential problem for in-operation health monitoring of civil structures is the variation of the environment itself. Unlike laboratory experiments, civil structure modal properties change during time as temperature and humidity vary. Traffic and comparable transient events also influence the structures. Thus, structural modal properties are modified by slow low variations, as well as fast transient non stationarities. From a damage detection point of view, the former has to be detected, whereas the latter has to be neglected and should not perturb the detection. Of course, from a structural health monitoring point of view the knowledge of the true load is itself of paramount importance.

In this context, the considered perturbations will be of two kinds, either

1/ the influence of the temperature on civil structures, such as bridges or wind energy converters : as we will notice, those induced variations can be modeled by a additive component on the system stiffness matrix depending on the current temperature, as

$$K = K_{struct} + K_T$$
.

We will then have to monitor the variations in  $K_{struct}$  independently of the variations in  $K_T$ , based on some measurements generated from a system, whose stiffness matrix is K.

2/ the influence of the aeroelastic forces on aeronautical structures such as aircrafts or rockets and on flexible civil structures such as long-span bridges : we will see as well that this influence implies a modification of the classical mechanical equation (2)

$$M\ddot{Z} + C\dot{Z} + KZ = V \tag{2}$$

where (M, C, K) are the mass, damping and stiffness matrices of the system and Z the associated vector of displacements measured on the monitored structure. In a first approximation, those quantities are related by (2). Assuming U is the velocity of the system, adding U dependent aeroelasticity terms, as in (3), introduces a coupling between U and (M, C, K).

$$M\ddot{Z} + C\dot{Z} + KZ = U^2DZ + UE\dot{Z} + V \tag{3}$$

Most of the research at Inria for a decade has been devoted to the study of subspace methods and how they handle the problems described above.

Model (2) is characterized by the following property (we formulate it for the single sensor case, to simplify notations): Let  $y_{-N} \cdots y_{+N}$  be the data set, where N is large, and let M, P sufficiently smaller than N for the following objects to make sense: 1/ define the row vectors  $Y_k = (y_k \cdots y_{k-M}), |k| \leq P$ ; 2/ stack the  $Y_k$ on top of each other for  $k = 0, 1, \dots, P$  to get the data matrix  $\mathcal{Y}_+$  and stack the column vectors  $Y_k^T$  for  $k = 0, -1, \dots, -P$  to get the data matrix  $\mathcal{Y}_-$ ; 3/ the product  $\mathcal{H} = \mathcal{Y}_+ \mathcal{Y}_-$  is a Hankel matrix. Then, matrix  $\mathcal{H}$  on the one hand, and the observability matrix  $\mathcal{O}(H, F)$  of system (2) on the other hand, possess almost identical left kernel spaces, asymptotically for M, N large. This property is the basis of subspace identification methods. Extracting  $\mathcal{O}(H, F)$  using some Singular Value Decomposition from  $\mathcal{H}$  then (H, F) from  $\mathcal{O}(H, F)$ using a Least Square approach has been the foundation of the academic work on subspace methods for many years. The team focused on the numerical efficiency and consistency of those methods and their applicability on solving the problems above. There are numerous ways to implement those methods. This approach has seen a wide acceptance in the industry and benefits from a large background in the automatic control literature. Up to now, there was a discrepancy between the a priori efficiency of the method and some not so efficient implementations of this algorithm. In practice, for the last ten years, stabilization diagrams have been used to handle the instability and the weakness with respect to noise, as well as the poor capability of those methods to determine model orders from data. Those methods implied some engineering expertise and heavy post processing to discriminate between models and noise. This complexity has led the mechanical community to adopt preferably frequency domain methods such as Polyreference LSCF. Our focus has been on improving the numerical stability of the subspace algorithms by studying how to compute the least square solution step in this algorithm. This yields to a very efficient noise free algorithm, which has provided a renewed acceptance in the mechanical engineering community for the subspace algorithms. Now we focus on improving speed and robustness of those algorithms.

Subspace methods can also be used to test whether a given data set conforms a model: just check whether this property holds, for a given pair {data, model}. Since equality holds only asymptotically, equality must be tested against some threshold  $\varepsilon$ ; tuning  $\varepsilon$  relies on so-called *asymptotic local* approach for testing between close hypotheses on long data sets — this method was introduced by Le Cam in the 70s. By using the Jacobian between pair (H, F) and the modes and mode shapes, or the Finite Element Model parameters, one can localize and assess the damage.

In oder to discriminate between damage and temperature variations, we need to monitor the variations in  $K_{struct}$  while being blind to the variations in  $K_T$ . In statistical terms, we must detect and diagnose changes in  $K_{struct}$  while rejecting nuisance parameter  $K_T$ . Several techniques were explored in the thesis of Houssein Nasser, from purely empirical approaches to (physical) model based approaches. Empirical approaches do work, but model based approaches are the most promising and constitue a focus of our future researches. This approach requires a physical model of how temperature affects stiffness in various materials. This is why a large part of our future research is devoted to the modeling of such environmental effect.

This approach has been used also for flutter monitoring in Rafik Zouari's PhD thesis for handling the aeroelastic effect.

# **3. Research Program**

### 3.1. Vibration analysis

In this section, the main features for the key monitoring issues, namely identification, detection, and diagnostics, are provided, and a particular instantiation relevant for vibration monitoring is described.

It should be stressed that the foundations for identification, detection, and diagnostics, are fairly general, if not generic. Handling high order linear dynamical systems, in connection with finite elements models, which call for using subspace-based methods, is specific to vibration-based SHM. Actually, one particular feature of model-based sensor information data processing as exercised in I4S, is the combined use of black-box or semi-physical models together with physical ones. Black-box and semi-physical models are, for example, eigenstructure parameterizations of linear MIMO systems, of interest for modal analysis and vibration-based SHM. Such models are intended to be identifiable. However, due to the large model orders that need to be considered, the issue of model order selection is really a challenge. Traditional advanced techniques from statistics such as the various forms of Akaike criteria (AIC, BIC, MDL, ...) do not work at all. This gives rise to new research activities specific to handling high order models.

Our approach to monitoring assumes that a model of the monitored system is available. This is a reasonable assumption, especially within the SHM areas. The main feature of our monitoring method is its intrinsic ability to the early warning of small deviations of a system with respect to a reference (safe) behavior under usual operating conditions, namely without any artificial excitation or other external action. Such a normal behavior is summarized in a reference parameter vector  $\theta_0$ , for example a collection of modes and mode-shapes.

### 3.1.1. Identification

The behavior of the monitored continuous system is assumed to be described by a parametric model  $\{\mathbf{P}_{\theta}, \theta \in \Theta\}$ , where the distribution of the observations  $(Z_0, ..., Z_N)$  is characterized by the parameter vector  $\theta \in \Theta$ .

For reasons closely related to the vibrations monitoring applications, we have been investigating subspacebased methods, for both the identification and the monitoring of the eigenstructure  $(\lambda, \phi_{\lambda})$  of the state transition matrix F of a linear dynamical state-space system :

$$\begin{cases} X_{k+1} = F X_k + V_{k+1} \\ Y_k = H X_k + W_k \end{cases},$$
(4)

namely the  $(\lambda, \varphi_{\lambda})$  defined by :

det 
$$(F - \lambda I) = 0$$
,  $(F - \lambda I) \phi_{\lambda} = 0$ ,  $\varphi_{\lambda} \stackrel{\Delta}{=} H \phi_{\lambda}$  (5)

The (canonical) parameter vector in that case is :

$$\theta \stackrel{\Delta}{=} \left( \begin{array}{c} \Lambda \\ \operatorname{vec} \Phi \end{array} \right) \tag{6}$$

where  $\Lambda$  is the vector whose elements are the eigenvalues  $\lambda$ ,  $\Phi$  is the matrix whose columns are the  $\varphi_{\lambda}$ 's, and vec is the column stacking operator.

Subspace-based methods is the generic name for linear systems identification algorithms based on either time domain measurements or output covariance matrices, in which different subspaces of Gaussian random vectors play a key role [62].

Let  $R_i \stackrel{\Delta}{=} \mathbf{E} \left( Y_k \; Y_{k-i}^T \right)$  and:

$$\mathcal{H}_{p+1,q} \triangleq \begin{pmatrix} R_1 & R_2 & \vdots & R_q \\ R_2 & R_3 & \vdots & R_{q+1} \\ \vdots & \vdots & \vdots & \vdots \\ R_{p+1} & R_{p+2} & \vdots & R_{p+q} \end{pmatrix} \triangleq \operatorname{Hank}(R_i)$$
(7)

be the output covariance and Hankel matrices, respectively; and:  $G \stackrel{\Delta}{=} \mathbf{E} (X_k Y_{k-1}^T)$ . Direct computations of the  $R_i$ 's from the equations (4) lead to the well known key factorizations :

$$R_{i} = HF^{i-1}G$$

$$\mathcal{H}_{p+1,q} = \mathcal{O}_{p+1}(H,F) \mathcal{C}_{q}(F,G)$$
(8)

where:

$$\mathcal{O}_{p+1}(H,F) \stackrel{\Delta}{=} \begin{pmatrix} H \\ HF \\ \vdots \\ HF^p \end{pmatrix} \quad \text{and} \quad \mathcal{C}_q(F,G) \stackrel{\Delta}{=} (G \ FG \ \cdots \ F^{q-1}G) \tag{9}$$

are the observability and controllability matrices, respectively. The observation matrix H is then found in the first block-row of the observability matrix  $\mathcal{O}$ . The state-transition matrix F is obtained from the shift invariance property of  $\mathcal{O}$ . The eigenstructure  $(\lambda, \phi_{\lambda})$  then results from (5).

Since the actual model order is generally not known, this procedure is run with increasing model orders.

### 3.1.2. Detection

Our approach to on-board detection is based on the so-called asymptotic statistical local approach. It is worth noticing that these investigations of ours have been initially motivated by a vibration monitoring application example. It should also be stressed that, as opposite to many monitoring approaches, our method does not require repeated identification for each newly collected data sample.

For achieving the early detection of small deviations with respect to the normal behavior, our approach generates, on the basis of the reference parameter vector  $\theta_0$  and a new data record, indicators which automatically perform :

- The early detection of a slight mismatch between the model and the data;
- A preliminary diagnostics and localization of the deviation(s);
- The tradeoff between the magnitude of the detected changes and the uncertainty resulting from the estimation error in the reference model and the measurement noise level.

These indicators are computationally cheap, and thus can be embedded. This is of particular interest in some applications, such as flutter monitoring.

Choosing the eigenvectors of matrix F as a basis for the state space of model (4) yields the following representation of the observability matrix:

$$\mathcal{O}_{p+1}(\theta) = \begin{pmatrix} \Phi \\ \Phi \Delta \\ \vdots \\ \Phi \Delta^p \end{pmatrix}$$
(10)

where  $\Delta \stackrel{\Delta}{=} \operatorname{diag}(\Lambda)$ , and  $\Lambda$  and  $\Phi$  are as in (6). Whether a nominal parameter  $\theta_0$  fits a given output covariance sequence  $(R_j)_i$  is characterized by:

$$\mathcal{O}_{p+1}(\theta_0)$$
 and  $\mathcal{H}_{p+1,q}$  have the same left kernel space. (11)

This property can be checked as follows. From the nominal  $\theta_0$ , compute  $\mathcal{O}_{p+1}(\theta_0)$  using (10), and perform e.g. a singular value decomposition (SVD) of  $\mathcal{O}_{p+1}(\theta_0)$  for extracting a matrix U such that:

$$U^T U = I_s \text{ and } U^T \mathcal{O}_{p+1}(\theta_0) = 0$$
 (12)

Matrix U is not unique (two such matrices relate through a post-multiplication with an orthonormal matrix), but can be regarded as a function of  $\theta_0$ . Then the characterization writes:

$$U(\theta_0)^T \mathcal{H}_{p+1,q} = 0 \tag{13}$$

#### 3.1.2.1. Residual associated with subspace identification.

Assume now that a reference  $\theta_0$  and a new sample  $Y_1, \dots, Y_N$  are available. For checking whether the data agree with  $\theta_0$ , the idea is to compute the empirical Hankel matrix  $\hat{\mathcal{H}}_{p+1,q}$ :

$$\widehat{\mathcal{H}}_{p+1,q} \stackrel{\Delta}{=} \operatorname{Hank}\left(\widehat{R}_{i}\right), \quad \widehat{R}_{i} \stackrel{\Delta}{=} 1/(N-i) \sum_{k=i+1}^{N} Y_{k} Y_{k-i}^{T}$$
(14)

and to define the residual vector:

$$\zeta_N(\theta_0) \stackrel{\Delta}{=} \sqrt{N} \operatorname{vec} \left( U(\theta_0)^T \, \widehat{\mathcal{H}}_{p+1,q} \right) \tag{15}$$

Let  $\theta$  be the actual parameter value for the system which generated the new data sample, and  $\mathbf{E}_{\theta}$  be the expectation when the actual system parameter is  $\theta$ . From (13), we know that  $\zeta_N(\theta_0)$  has zero mean when no change occurs in  $\theta$ , and nonzero mean if a change occurs. Thus  $\zeta_N(\theta_0)$  plays the role of a residual.

As in most fault detection approaches, the key issue is to design a *residual*, which is ideally close to zero under normal operation, and has low sensitivity to noises and other nuisance perturbations, but high sensitivity to small deviations, before they develop into events to be avoided (damages, faults, ...). The originality of our approach is to :

- Design the residual basically as a parameter estimating function,
- *Evaluate* the residual thanks to a kind of central limit theorem, stating that the residual is asymptotically Gaussian and reflects the presence of a deviation in the parameter vector through a change in its own mean vector, which switches from zero in the reference situation to a non-zero value.

The central limit theorem shows [56] that the residual is asymptotically Gaussian :

$$\zeta_{N} \xrightarrow{N \to \infty} \begin{cases} & \mathcal{N}(0, \Sigma) & \text{under } \mathbf{P}_{\theta_{0}} &, \\ & & \\ & & \\ & & \mathcal{N}(\mathcal{J}\eta, \Sigma) & \text{under } \mathbf{P}_{\theta_{0} + \eta/\sqrt{N}} &, \end{cases}$$
(16)

where the asymptotic covariance matrix  $\Sigma$  can be estimated, and manifests the deviation in the parameter vector by a change in its own mean value. Then, deciding between  $\eta = 0$  and  $\eta \neq 0$  amounts to compute the following  $\chi^2$ -test, provided that  $\mathcal{J}$  is full rank and  $\Sigma$  is invertible :

$$\chi^2 = \overline{\zeta}^T \mathbf{F}^{-1} \overline{\zeta} \gtrless \lambda .$$
 (17)

where

$$\overline{\zeta} \stackrel{\Delta}{=} \mathcal{J}^T \Sigma^{-1} \zeta_N \text{ and } \mathbf{F} \stackrel{\Delta}{=} \mathcal{J}^T \Sigma^{-1} \mathcal{J}$$
 (18)

### 3.1.3. Diagnostics

A further monitoring step, often called *fault isolation*, consists in determining which (subsets of) components of the parameter vector  $\theta$  have been affected by the change. Solutions for that are now described. How this relates to diagnostics is addressed afterwards.

The question: which (subsets of) components of  $\theta$  have changed ?, can be addressed using either nuisance parameters elimination methods or a multiple hypotheses testing approach [55].

In most SHM applications, a complex physical system, characterized by a generally non identifiable parameter vector  $\Phi$  has to be monitored using a simple (black-box) model characterized by an identifiable parameter vector  $\theta$ . A typical example is the vibration monitoring problem for which complex finite elements models are often available but not identifiable, whereas the small number of existing sensors calls for identifying only simplified input-output (black-box) representations. In such a situation, two different diagnosis problems may arise, namely diagnosis in terms of the black-box parameter  $\theta$  and diagnosis in terms of the parameter vector  $\Phi$  of the underlying physical model.

The isolation methods sketched above are possible solutions to the former. Our approach to the latter diagnosis problem is basically a detection approach again, and not a (generally ill-posed) inverse problem estimation approach.

The basic idea is to note that the physical sensitivity matrix writes  $\mathcal{J}\mathcal{J}_{\Phi\theta}$ , where  $\mathcal{J}_{\Phi\theta}$  is the Jacobian matrix at  $\Phi_0$  of the application  $\Phi \mapsto \theta(\Phi)$ , and to use the sensitivity test for the components of the parameter vector  $\Phi$ . Typically this results in the following type of directional test :

$$\chi_{\Phi}^{2} = \zeta^{T} \Sigma^{-1} \mathcal{J} \mathcal{J}_{\Phi\theta} \left( \mathcal{J}_{\Phi\theta}^{T} \mathcal{J}^{T} \Sigma^{-1} \mathcal{J} \mathcal{J}_{\Phi\theta} \right)^{-1} \mathcal{J}_{\Phi\theta}^{T} \mathcal{J}^{T} \Sigma^{-1} \zeta \gtrless \lambda .$$
<sup>(19)</sup>

It should be clear that the selection of a particular parameterization  $\Phi$  for the physical model may have a non negligible influence on such type of tests, according to the numerical conditioning of the Jacobian matrices  $\mathcal{J}_{\Phi\theta}$ .

# **3.2.** Thermal methods

### 3.2.1. Infrared thermography and heat transfer

This section introduce the infrared radiation and its link with the temperature, in the next part different measurement methods based on that principle are presented.

#### 3.2.1.1. Infrared radiation

Infrared is an electromagnetic radiation having a wavelength between  $0.2\mu m$  and 1 mm, this range begin in uv spectrum and it ends on the microwaves domain, see Figure 1.



Figure 1. Electromagnetic spectrum - Credit MODEST, M.F. (1993). Radiative Heat Transfer. Academic Press.

Band name	wavelength	Uses \ definition
Near infrared (PIR, IR-A, NIR)	$0.7 - 3\mu m$	Reflected solar heat flux
Mid infrared (MIR, IR-B)	$3-50\mu m$	Thermal infrared
Far infrared (LIR, IR-C, FIR)	$50 - 1000 \mu m$	Astronomy

Table 1. Wavelength bands in the infrared according to ISO 20473:2007

For scientific purpose infrared can be divided in three ranges of wavelength in which the application varies, see Table 1.

Our work is concentrated in the mid infrared spectral band. Keep in mind that Table 1 represents the ISO 20473 division scheme, in the literature boundaries between bands can move slightly.

The Plank's law, proposed by Max Planck in 1901, allows to compute the black body emission spectrum for various temperatures (and only temperatures), see Figure 2 left. The black body is a theoretical construction, it represents perfect energy emitter at a given temperature, cf Equation (20).

$$M^{o}_{\lambda,T} = \frac{C_1 \lambda^{-5}}{\exp^{\frac{C_2}{\lambda T}} - 1}$$
(20)

With  $\lambda$  the wavelength in m and T as the temperature in Kelvin. The  $C_1$  an  $C_2$  constant, respectively in W.m<sup>2</sup> and m.K are defined as follow:

$$C_1 = 2hc^2\pi$$

$$C_2 = h\frac{c}{k}$$
(21)

with

- c The electromagnetic wave speed (in vacuum c is the light speed in m.s<sup>-1</sup>).
- $k = 1.381e^{-23}$  J.K<sup>-1</sup> The Boltzmann (Entropy definition from Ludwig Boltzmann 1873). It can be seen as a proportionality factor between the temperature and the energy of a system.
- $h \approx 6,62606957 e^{-34}$  J.s The Plank constant. It is the link between the photons energy and their frequency.



Figure 2. Left: Plank's law at various temperatures - Right: Energy spectrum of the atmosphere

By generalizing the Plank's law with the Stefan Boltzmann law (proposed first in 1879 and then in 1884 by Joseph Stefan and Ludwig Boltzmann) it is possible to address mathematically the energy spectrum of real body at each wavelength dependent of the temperature, the optical condition and the real body properties, which is the base of the infrared thermography.

For example, Figure 2 right presents the energy spectrum of the atmosphere at various levels, it can be seen that the various properties of the atmosphere affect the spectrum at various wavelengths. Other important point is that the infrared solar heat flux can be approximated by a black body at 5523,15 K.

### 3.2.1.2. Infrared Thermography

The infrared thermography is a way to measure the thermal radiation received from a medium. With that information about the electromagnetic flux it is possible to estimate the surface temperature of the body, see section 3.2.1.1. Various types of detector can assure the measure of the electromagnetic radiation.

Those different detectors can take various forms and/or manufacturing process. For our research purpose we use uncooled infrared camera using a matrix of microbolometers detectors. A microbolometer, as a lot of transducers, converts a radiation in electric current used to represent the physical quantity (here the heat flux).

This field of activity includes the use and the improvement of vision system, like in [3].

### 3.2.2. Heat transfer theory

Once the acquisition process is done, it is useful to model the heat conduction inside the cartesian domain  $\Omega$ . Note that in opaque solid medium the heat conduction is the only mode of heat transfer. Proposed by Jean Baptiste Biot in 1804 and experimentally demonstrated by Joseph Fourier in 1821, the Fourier Law describes the heat flux inside a solid, cf Equation (22).

$$\varphi = k\nabla T \quad X \in \Omega \tag{22}$$

Where k is the thermal conductivity in W.m<sup>-1</sup>.K °,  $\nabla$  is the gradient operator and  $\varphi$  is the heat flux density in Wm<sup>-2</sup>. This law illustrates the first principle of thermodynamic (law of conservation of energy) and implies the second principle (irreversibility of the phenomenon), from this law it can be seen that the heat flux always goes from hot area to cold area.

An energy balance with respect to the first principle drives to the expression of the heat conduction in all point of the domain  $\Omega$ , cf Equation (23). This equation has been proposed by Joseph Fourier in 1811.

$$\rho C \frac{\partial T(X,t)}{\partial t} = \nabla \cdot (k \nabla T) + P \quad X \in \Omega$$
<sup>(23)</sup>

With  $\nabla$ .() the divergence operator, C the specific heat capacity in J.kg<sup>-1</sup>.  ${}^{o}$ K<sup>-1</sup>,  $\rho$  the volumetric mass density in kg. m<sup>-3</sup>, X the space variable  $X = \{x, y, z\}$  and P a possible internal heat production in W.m<sup>-3</sup>. To solve the system (23), it is necessary to express the boundaries conditions of the system. With the developments presented in section 3.2.1.1 and the Fourier's law it is possible, for example, to express the thermal radiation and the convection phenomenon which can occur at  $\partial\Omega$  the system boundaries, cf Equation (24).

$$\varphi = k\nabla T \cdot n = \underbrace{h\left(T_{fluid} - T_{Boundarie}\right)}_{\text{Convection}} + \underbrace{\epsilon\sigma_s\left(T_{environement}^4 - T_{Boundary}^4\right)}_{\text{Radiation}} + \varphi_0 \quad X \in \partial\Omega \tag{24}$$

Equation (24) is the so called Robin condition on the boundary  $\partial\Omega$ , where *n* is the normal, *h* the convective heat transfer coefficient in W.m<sup>-2</sup>.K<sup>-1</sup> and  $\varphi_0$  an external energy contribution W.m<sup>-2</sup>, in cases where the external energy contribution is artificial and controlled we call it active thermography (spotlight etc...) in the contrary it is called passive thermography (direct solar heat flux).

The systems presented in the different sections above (3.2.1 to 3.2.2) are useful to build physical models in order to represents the measured quantity. To estimate key parameters, as the conductivity, one way to do is the model inversion, the next section will introduce that principle.

### 3.2.3. Inverse model for parameters estimation

Lets take any model A which can for example represent the conductive heat transfer in a medium, the model is solved for a parameter vector P and it results another vector b, cf Equation (25). For example if A represents the heat transfer, b can be the temperature evolution.

$$AP = b \tag{25}$$

With A a matrix of size  $n \times m$ , P a vector of size m and b of size n, preferentially  $n \gg P$ . This model is called direct model, the inverse model consist to find a vector P which satisfy the results b of the direct model. For that we need to inverse the matrix A, cf Equation (26).

$$P = A^{-1}b \tag{26}$$

Here we want find the solution AP which is closest to the acquired measures M, Equation (27).

$$AP \approx \mathcal{M}$$
 (27)

To do that it is important to respect the well posed condition established by Jacques Hadamard in 1902

- A solution exists.
- The solution is unique.
- The solution's behavior changes continuously with the initial conditions.

Unfortunately those condition are rarely respected in our field of study. That is why we dont solve directly the system (27) but we minimise the quadratic coast function (28) which represents the Legendre-Gauss least square algorithm for linear problems.

$$min_{P}\left(\left\|AP - \mathcal{M}\right\|^{2}\right) = min_{P}\left(\mathfrak{F}\right)$$
(28)

Where  $\mathcal{F}$  can be a product of matrix.

$$\mathcal{F} = \left[AP - \mathcal{M}\right]^T \left[AP - \mathcal{M}\right]$$

In some case the problem is still ill-posed and need to be regularized for example using the Tikhonov regularization. An elegant way to minimize the cost function  $\mathcal{F}$  is compute the gradient, Equation (29) and find where it is equal to zero.

$$\nabla \mathcal{F}(P) = 2 \left[ -\frac{\partial A P^T}{\partial P} \right] \left[ A P - \mathcal{M} \right] = 2J(P)^T \left[ A P - \mathcal{M} \right]$$
(29)

Where J is the sensitivity matrix of the model A to its parameter vector P.

Until now the inverse method proposed is valid only when the model A is linearly dependent of its parameter P, for the heat equation it is the case when you want to estimate the external heat flux,  $\varphi_0$  in equation 24. For all the other parameters, like the conductivity k the model is non-linearly dependant of its parameter P. For such case the use of iterative algorithm is needed, for example the Levenberg-Marquardt algorithm, cf Equation (30).

$$P^{k+1} = P^{k} + \left[ \left( J^{k} \right)^{T} J^{k} + \mu^{k} \Omega^{k} \right]^{-1} \left( J^{k} \right)^{T} \left[ \mathcal{M} - A(P^{k}) \right]$$
(30)

Equation (30) is solved iteratively at each loop k. Some of our results with such linear or non linear method can be seen in [4] or [2], more specifically [1] is a custom implementation of the Levenberg-Marquardt algorithm based on the adjoint method (developed by Jacques Louis Lions in 1968) coupled to the conjugate gradient algorithm to estimate wide properties field in a medium.

# **3.3. Reflectometry-based methods for electrical engineering and for civil engineering**

The fast development of electronic devices in modern engineering systems involves more and more connections through cables, and consequently, with an increasing number of connection failures. Wires and connectors are subject to ageing and degradation, sometimes under severe environmental conditions. In many applications, the reliability of electrical connexions is related to the quality of production or service, whereas in critical applications reliability becomes also a safety issue. It is thus important to design smart diagnosis systems able to detect connection defects in real time. This fact has motivated research projects on methods for fault diagnosis in this field. Some of these projects are based on techniques of reflectometry, which consist in injecting waves into a cable or a network and in analyzing the reflections. Depending on the injected waveforms and on the methods of analysis, various techniques of reflectometry are available. They all have the common advantage of being non destructive.

At Inria the research activities on reflectometry started within the SISYPHE EPI several years ago and now continue in the I4S EPI. Our most notable contribution in this area is a method based on the *inverse scattering* theory for the computation of *distributed characteristic impedance* along a cable from reflectometry measurements [14], [11], [61]. It provides an efficient solution for the diagnosis of *soft* faults in electrical cables, like in the example illustrated in Figure 3. While most reflectometry methods for fault diagnosis are based on the detection and localization of impedance discontinuity, our method yielding the spatial profile of the characteristic impedance is particularly suitable for the diagnosis of soft faults *with no or weak impedance discontinuities*.

Fault diagnosis for wired networks have also been studied in Inria [63], [59]. The main results concern, on the one hand, simple star-shaped networks from measurements made at a single node, on the other hand, complex networks of arbitrary topological structure with complete node observations.

Though initially our studies on reflectometry were aiming at applications in electrical engineering, since the creation of the I4S team, we are also investigating applications in the field of civil engineering, by using electrical cables as sensors for monitoring changes in mechanical structures.

What follows is about some basic elements on mathematical equations of electric cables and networks, the main approach we follow in our study, and our future research directions.

### 3.3.1. Mathematical model of electric cables and networks

A cable excited by a signal generator can be characterized by the telegrapher's equations [60]

$$\frac{\partial}{\partial z}V(t,z) + L(z)\frac{\partial}{\partial t}I(t,z) + R(z)I(t,z) = 0$$

$$\frac{\partial}{\partial z}I(t,z) + C(z)\frac{\partial}{\partial t}V(t,z) + G(z)V(t,z) = 0$$
(31)



Figure 3. Inverse scattering software (ISTL) for cable soft fault diagnosis.

where t represents the time, z is the longitudinal coordinate along the cable, V(t, z) and I(t, z) are respectively the voltage and the current in the cable at the time instant t and at the position z, R(z), L(z), C(z) and G(z)denote respectively the series resistance, the inductance, the capacitance and the shunt conductance per unit length of the cable at the position z. The left end of the cable (corresponding to z = a) is connected to a voltage source  $V_s(t)$  with internal impedance  $R_s$ . The quantities  $V_s(t)$ ,  $R_s$ , V(t, a) and I(t, a) are related by

$$V(t,a) = V_s(t) - R_s I(t,a).$$
(32)

At the right end of the cable (corresponding to z = b), the cable is connected to a load of impedance  $R_L$ , such that

$$V(t,b) = R_L I(t,b).$$
(33)

One way for deriving the above model is to spatially discretize the cable and to characterize each small segment with 4 basic lumped parameter elements for the *j*-th segment: a resistance  $\Delta R_j$ , an inductance  $\Delta L_j$ , a capacitance  $\Delta C_j$  and a conductance  $\Delta G_j$ . The entire circuit is described by a system of ordinary differential equations. When the spatial discretization step size tends to zero, the limiting model leads to the telegrapher's equations.

A wired network is a set of cables connected at some nodes, where loads and sources can also be connected. Within each cable the current and voltage satisfy the telegrapher's equations, whereas at each node the current and voltage satisfy the Kirchhoff's laws, unless in case of connector failures.

### 3.3.2. The inverse scattering theory applied to cables

The inverse scattering transform was developed during the 1970s-1980s for the analysis of some nonlinear partial differential equations [58]. The visionary idea of applying this theory to solving the cable inverse problem goes also back to the 1980s [57]. After having completed some theoretic results directly linked to practice [14], [61], we started to successfully apply the inverse scattering theory to cable soft fault diagnosis, in collaboration with GEEPS-SUPELEC [11].

To link electric cables to the inverse scattering theory, the telegrapher's equations are transformed in a few steps to fit into a particular form studied in the inverse scattering theory. The Fourier transform is first applied to obtain a frequency domain model, the spatial coordinate z is then replaced by the propagation time

$$x(z) = \int_0^z \sqrt{L(s)C(s)} ds$$

and the frequency domain variables  $V(\omega, x), I(\omega, x)$  are replaced by the pair

$$\nu_{1}(\omega, x) = \frac{1}{2} \left[ Z_{0}^{-\frac{1}{2}}(x)U(\omega, x) - Z_{0}^{\frac{1}{2}}(x)I(\omega, x) \right]$$

$$\nu_{2}(\omega, x) = \frac{1}{2} \left[ Z_{0}^{-\frac{1}{2}}(x)U(\omega, x) + Z_{0}^{\frac{1}{2}}(x)I(\omega, x) \right]$$
(34)

with

$$Z_0(x) = \sqrt{\frac{L(x)}{C(x)}}.$$
(35)

These transformations lead to the Zakharov-Shabat equations

$$\frac{d\nu_1(\omega, x)}{dx} + ik\nu_1(\omega, x) = q^*(x)\nu_1(\omega, x) + q^+(x)\nu_2(\omega, x)$$

$$\frac{d\nu_2(\omega, x)}{dx} - ik\nu_2(\omega, x) = q^-(x)\nu_1(\omega, x) - q^*(x)\nu_2(\omega, x)$$
(36)

with

$$q^{\pm}(x) = -\frac{1}{4} \frac{d}{dx} \left[ ln \frac{L(x)}{C(x)} \right] \mp \frac{1}{2} \left[ \frac{R(x)}{L(x)} - \frac{G(x)}{C(x)} \right] = -\frac{1}{2Z_0(x)} \frac{d}{dx} Z_0(x) \mp \frac{1}{2} \left[ \frac{R(x)}{L(x)} - \frac{G(x)}{C(x)} \right] q^*(x) = \frac{1}{2} \left[ \frac{R(x)}{L(x)} + \frac{G(x)}{C(x)} \right].$$
(37)

These equations have been well studied in the inverse scattering theory, for the purpose of determining partly the "potential functions"  $q^{\pm}(x)$  and  $q^{*}(x)$  from the scattering data matrix, which turns out to correspond to the data typically collected with reflectometry instruments. For instance, it is possible to compute the function  $Z_0(x)$  defined in (35), often known as the characteristic impedance, from the reflection coefficient measured at one end of the cable. Such an example is illustrated in Figure 3. Any fault affecting the characteristic impedance, like in the example of Figure 3 caused by a slight geometric deformation, can thus be efficiently detected, localized and characterized.

# 3.4. Research Program

The research will first focus on the extension and implementation of current techniques as developed in I4S and IFSTTAR. Before doing any temperature rejection on large scale structures as planned, we need to develop good and accurate models of thermal fields. We also need to develop robust and efficient versions of our algorithms, mainly the subspace algorithms before envisioning linking them with physical models. Briefly, we need to mature our statistical toolset as well as our physical modeling before mixing them together later on.

### 3.4.1. Vibration analysis and monitoring

#### 3.4.1.1. Direct vibration modeling under temperature changes

This task builds upon what has been achieved in the CONSTRUCTIF project, where a simple formulation of the temperature effect has been exhibited, based on relatively simple assumptions. The next step is to generalize this modeling to a realistic large structure under complex thermal changes. Practically, temperature and resulting structural prestress and pre strains of thermal origin are not uniform and civil structures are complex. This leads to a fully 3D temperature field, not just a single value. Inertia effects also forbid a trivial prediction of the temperature based on current sensor outputs while ignoring past data. On the other side, the temperature is seen as a nuisance. That implies that any damage detection procedure has first to correct the temperature effect prior to any detection.

Modeling vibrations of structures under thermal prestress does and will play an important role in the static correction of kinematic measurements, in health monitoring methods based on vibration analysis as well as in durability and in the active or semi-active control of civil structures that by nature are operated under changing environmental conditions. As a matter of fact, using temperature and dynamic models the project aims at correcting the current vibration state from induced temperature effects, such that damage detection algorithms rely on a comparison of this thermally corrected current vibration state with a reference state computed or measured at a reference temperature. This approach is expected to cure damage detection algorithms from the environmental variations.

I4S will explore various ways of implementing this concept, notably within the FUI SIPRIS project.

### 3.4.1.2. Damage localization algorithms (in the case of localized damages such as cracks)

During the CONSTRUCTIF project, both feasibility and efficiency of some damage detection and localization algorithms were proved. Those methods are based on the tight coupling of statistical algorithms with finite element models. It has been shown that effective localization of some damaged elements was possible, and this was validated on a numerical simulated bridge deck model. Still, this approach has to be validated on real structures.

On the other side, new localization algorithms are currently investigated such as the one developed conjointly with University of Boston and tested within the framework of FP7 ISMS project. These algorithms will be implemented and tested on the PEGASE platform as well as all our toolset.

When possible, link with temperature rejection will be done along the lines of what has been achieved in the CONSTRUCTIF project.

### 3.4.1.3. Uncertainty quantification for system identification algorithms

Some emphasis will be put on expressing confidence intervals for system identification. It is a primary goal to take into account the uncertainty within the identification procedure, using either identification algorithms derivations or damage detection principles. Such algorithms are critical for both civil and aeronautical structures monitoring. It has been shown that confidence intervals for estimation parameters can theoretically be related to the damage detection techniques and should be computed as a function of the Fisher information matrix associated to the damage detection test. Based on those assumptions, it should be possible to obtain confidence intervals for a large class of estimates, from damping to finite elements models. Uncertainty considerations are also deeply investigated in collaboration with Dassault Aviation in Mellinger PhD thesis or with Northeastern University, Boston, within Gallegos PhD thesis.

### 3.4.2. Reflectometry-based methods for civil engineering structure health monitoring

The inverse scattering method we developed is efficient for the diagnosis of all soft faults affecting the characteristic impedance, the major parameter of a cable. In some particular applications, however, faults would rather affect the series resistance (ohmic loss) or shunt conductance (leakage loss) than the characteristic impedance. The first method we developed for the diagnosis of such losses had some numerical stability problems. The new method is much more reliable and efficient. It is also important to develop efficient solutions for long cables, up to a few kilometers.

For wired networks, the methods we already developed cover either the case of simple networks with a single node measurement or the case of complex networks with complete node measurements. Further developments are still necessary for intermediate situations.

In terms of applications, the use of electric cables as sensors for the monitoring of various structures is still at its beginning. We believe that this new technology has a strong potential in different fields, notably in civil engineering and in materials engineering.

### 3.4.3. Non Destructive testing of CFRP bonded on concrete through active thermography

Strengthening or retrofitting of reinforced concrete structures by externally bonded fibre-reinforced polymer (FRP) systems is now a commonly accepted and widespread technique. However, the use of bonding techniques always implies following rigorous installation procedures. The number of carbon fibre-reinforced polymer (CFRP) sheets and the glue layer thickness are designed by civil engineers to address strengthening objectives. Moreover, professional crews have to be trained accordingly in order to ensure the durability and long-term performance of the FRP reinforcements. Conformity checking through an 'in situ' verification of the bonded FRP systems is then highly desirable. The quality control programme should involve a set of adequate inspections and tests. Visual inspection and acoustic sounding (hammer tap) are commonly used to detect delaminations (disbonds). Nevertheless, these techniques are unable to provide sufficient information about the depth (in case of multilayered composite) and width of the disbonded areas. They are also incapable of evaluating the degree of adhesion between the FRP and the substrate (partial delamination, damage of the resin and poor mechanical properties of the resin). Consequently, rapid and efficient inspection methods are required. Among the non-destructive (NDT) methods currently under study, active infrared thermography is investigated due to its ability to be used in the field. In such context and to reach the aim of having an in situ efficient NDT method, we carried out experiments and subsequent data analysis using thermal excitation. Image processing, inverse thermal modelling and 3D numerical simulations are used and then applied to experimental data obtained in laboratory conditions.

### 3.4.4. IRSHM: Multi-Sensing system for outdoor thermal monitoring

Ageing of transport infrastructures combined with traffic and climatic solicitations contribute to the reduction of their performances. To address and quantify the resilience of civil engineering structure, investigations on robust, fast and efficient methods are required. Among research works carried out at IFSTTAR, methods for long term monitoring face an increasing demand. Such works take benefits of this last decade technological progresses in ICT domain.

Thanks to IFSTTAR years of experience in large scale civil engineering experiment, I4S is able to perform very long term thermal monitoring of structures exposed to environmental condition, as the solar heat flux, natural convection or seasonal perturbation. Informations system are developed to asses the data acquisition and researchers work on the quantification of the data to detect flaws emergence on structure, those techniques are also used to diagnose thermal insulation of buildings or monitoring of guided transport infrastructures, Figure 4 left. Experiments are carried out on a real transport infrastructure open to traffic and buildings. The detection of the inner structure of the deck is achieved by image processing techniques (as FFT), principal component thermography (PCT), Figure 4 right, or characterization of the inner structure thanks to an original image processing approach.

For the next few years, I4S is actively implied in the SenseCity EQUIPEX (http://sense-city.ifsttar.fr/) where our informations systems are used to monitor a mini-city replica, Figure 5.

### 3.4.5. R5G: The 5th Generation Road

The road has to reinvent itself periodically in response to innovations, societal issues and rising user expectations. The 5th Generation Road (R5G) focuses firmly on the future and sets out to be automated, safe, sustainable and suited to travel needs. Several research teams are involved in work related to this flagship project for IFSTTAR, which is a stakeholder in the Forever Open Road. Through its partnership with the COSYS (IFSTTAR) department, I4S is fully implicated in the development of the 5th Generation Road.



Figure 4. Left: Image in the visible spectrum of the deck surface - Right: PCT result on a bridge deck



Figure 5. Various view and results of the SenseCity experimentation site - (site and hardware view, IR imaging, Environmental Monitoring)

Most of the innovations featured in R5G are now mature, for example communication and few solutions for energy exchange between the infrastructure, the vehicle and the network manager; recyclable materials with the potential for self-diagnosis and repair, a pavement surface that remains permanently optimal irrespective of climatic variations... Nevertheless, implementing them on an industrial scale at a reasonable cost still represents a real challenge. Consultation with the stakeholders (researchers, industry, road network owners and users) has already established the priorities for the creation of full-scale demonstrators. The next stages are to achieve synergy between the technologies tested by the demonstrators, to manage the interfaces and get society to adopt R5G.

# 4. Application Domains

# 4.1. Civil Engineering

For at least three decades, monitoring the integrity of the civil infrastructure has been an active research topic because of major economical and societal issues, such as durability and safety of infrastructures, buildings and networks. Control of civil structures began a century ago. At stake is the mastering of the ageing of the bridges, as in America (US, Canada) and Great Britain, or the resistance to seismic events and the protection of the cultural heritage, as in Italy and Greece. The research effort in France is very ancient since for example early developments of optical methods to monitor civil structures began in the 70s and SHM practice can be traced back to the 50s with the vibrating wire sensors as strain gauges for dams. Stille the number of sensors actually placed on civil structures is kept to a minimum, mainly for cost reasons, but also because the return on investment sensing and data processing technologies is not properly established for civil structures. One of the current thematic priorities of the C2D2 governmental initiative is devoted to construction monitoring and

diagnostics. The picture in Asia (Japan, and also China) is somewhat different, in that recent or currently built bridges are equipped with hundreds if not thousands of sensors, in particular the Hong Kong-Shenzen Western Corridor and Stonecutter Bridge projects. However, the actual use of available data for operational purpose remains unclear.

Among the challenges for vibration-based bridges health monitoring, two major issues are the different kinds of (non measured) excitation sources and the environmental effects. Typically the traffic on *and* under the bridge, the wind and also the rain, contribute to excite the structure, and influence the measured dynamics. Moreover, the temperature is also known to affect the eigenfrequencies and mode-shapes, to an extent which can be significant w.r.t. the deviations to be monitored.

Thermomechanical prestress states affect the dynamic and the static behavior of most bridges, not only of very long and flexible ones. So, the reliable and fast determination of the state of prestress and prestrain associated with a temperature field becomes a crucial step in several engineering processes such as the health monitoring of civil structures. The best possible reconstruction of the temperature field could then become part of a complete process including massively distributed sensing of thermomechanical information on the structure, modeling and algorithms for the on-line detection of damages in the sense of abnormalities with regard to a nominal state, the whole chain being encapsulated in professional tools used by engineers in charge of real-life structural monitoring. For lack of an adequate mobilization of the useful multidisciplinary skills, this way remains about unexplored today.

# 4.2. Electrical cable and network monitoring

The fast development of electronic devices in modern engineering systems comes with more and more connections through cables, and consequently, the reliability of electric connections becomes a crucial issue. For example, in a modern automotive vehicle, the total length of onboard cables has tremendously increased during the last decades and is now up to 4km. These wires and connectors are subject to ageing or degradation because of severe environmental conditions. In this area, reliability becomes a safety issue. In some other domains, cable defects may have catastrophic consequences. It is thus a crucial challenge to design smart embedded diagnosis systems able to detect wired connection defects in real time. This fact has motivated research projects on methods for fault diagnosis in electric transmission lines and wired networks. Original methods have been recently developed by Inria, notably based on the inverse scattering theory, for cable and network monitoring. Further developments concern both theoretic study and industrial applications.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

The Structual Health Monitoring system developed by Vincent Le Cam and SDEL-CC for lightning detection and localization on electrical lines, has received the VINCI Innovation Award for Western France 2017. https://team.inria.fr/i4s/vinci-2017-innovation-award/

## 5.1.1. Awards

BEST PAPERS AWARDS :
[29] SPIE - Thermosense: Thermal Infrared Applications XXXIX. J. DUMOULIN, A. CRINIÈRE.
[30] IOMAC - 7th International Operational Modal Analysis Conference. M. Döhler, P. Andersen, L. MEVEL.

# 6. New Software and Platforms

## 6.1. PEGASE

Plate-forme Experte Générique pour Applications Sans-fil Embarquées

### KEYWORD: SHM (Structural Health Monitoring)

FUNCTIONAL DESCRIPTION: PEGASE is a generic high level wireless sensor platform. The main characteristics of PEGASE to reach this genericity are obtained by :

- Software genericity: use of a Linux embedded OS to make any developed application independent from the hardware, and to enable the user to manage the system without any physical operations.
- Hardware genericity: with a principle of daughter and mother boards, each redundant need is embedded (processing, memory, timing, GPS, energy, etc) where each pluggable daughter board implements a specific function (e.g., sensing, 3G, Lora/Sigfox and Ethernet wireless communications, signal processing and relay control).
- Accurate time synchronization: based on an original GPS and PPS algorithm, PEGASE platform is one of few boards able to time-stamp data from sensors or any event with an accuracy of some micro-seconds Universal Time.

After the industrial exploitation of PEGASE 1 (hundreds are sold), PEGASE 2 and the future PEGASE 3 version maintain and extend the previous platform. Focus on main characteristics is subject of electronic research and development:

- embed a "Debian" Linux operating system able to be validated for critical applications (such as SHM applications)
- embed a module dedicated to energy autonomy, to harvest energy from solar cells while considering the dis/charge of Lithium battery
- integrate a 3D accelerometer based on a MEMs to propose motion applications (train detection by vibration for example)
- new original daughter boards for new wireless IOT industrial protocols: LorA and Sigfox
- convert the proposed SDK (Single Development Kit) fully from C to C++ language
- a generic embedded front-end development called Zeus able to manage time control of Linux enslaved to the UTC synchronization, applications manager, network manager (from WiFi, Lora to 3G...), ...

Since 2017, PEGASE 2 platform is also used as the support for some lectures given at University of Nantes.

Associated to PEGASE hardware platform, I4S has also designed and programmed a generic Cloud Supervisor that allows to manage wireless sensors. In 2017 this application has matured, and has been licensed to two companies for industrial exploitation and distribution (Stimio and Power-Lan).

- Participants: Laurent Mevel, Mathieu Le Pen, Michael Doehler and Vincent Le Cam
- Contact: Michael Doehler
- URL: http://www.a3ip.com/joomla/index.php?option=com\_content&view=article&id=12&Itemid=8

# 6.2. TrackingMecaSysEvo

KEYWORDS: Particular filter - Kalman filter - Monte-Clarlo - Bayesian estimation - Vibrating system FUNCTIONAL DESCRIPTION: Based on a IPKF (Interacting Particles and Kalman Filter) implementation, TrackingMecaSysEvo allow mechanical parameters traking over the time for a 1-2-3D vibrating model. The algorithme insure also, input force and ambiant noise estimation

- Participants: Antoine Crinière, Laurent Mevel and Subhamoy Sen
- Partner: IFSTTAR
- Contact: Laurent Mevel

# 7. New Results

# 7.1. Outdoor InfraRed Thermography

### 7.1.1. Joint thermal and electromagnetic diagnostics

Participants: Nicolas Le Touz, Jean Dumoulin.

In this study, we present an inversion approach to detect and localize inclusions in thick walls under natural solicitations. The approach is based on a preliminary analysis of surface temperature field evolution with time (for instance acquired by infrared thermography); subsequently, this analysis is improved by taking advantage of a priori information provided by ground penetrating radar reconstruction of the structure under investigation. In this way, it is possible to improve the accuracy of the images achievable with the standalone thermal reconstruction method in the case of quasiperiodic natural excitation. [19]

### 7.1.2. Long term monitoring of transport infrastructures: from deployment to standardization Participants: Antoine Crinière, Jean Dumoulin, Laurent Mevel.

Long term monitoring of transport infrastructures by infrared thermography has been studied and tested on different structures. A first standalone infrared system architecture developed is presented and discussed. Results obtained with such system on different Civil Engineering structures are presented. Some data processing approaches and inverse thermal model for data analysis are introduced and discussed. Lessons learned from experiments carried out in outdoor with such system are listed and analyzed. Then, a new generation of infrared system architecture is proposed. Finally, conclusions and perspectives are addressed.[29], [46]

### 7.1.3. Infrared data reconstruction and calibration for long term monitoring

Participants: Thibaud Toullier, Jean Dumoulin, Laurent Mevel.

This study focuses on the evaluation and improvement of thermal instrumentation solutions for long-term monitoring of next-generation transport infrastructure. A test site was equipped with thermocouples and an infrared thermography system coupled with monitoring of environmental parameters. A method of spatial reconstruction of infrared images is presented. Measurement data acquired on site and then post-processed are analysed over time. A conclusion on the results achieved and prospects are proposed [48]

## 7.2. Data management of Smart territories and cities

Participants: Antoine Crinière, Jean Dumoulin.

Highly instrumented Smart-cities, which are now a common urban policies, are facing problems of management and storage of a large volume of data coming from an increasing number of sources. This study presents a data compression method by predictive coding of spatially correlated multi-source data based on reference selection and prediction by Kriging [47]

# 7.3. Smarts roads and R5G

# 7.3.1. Energy exchange modelization and infrared monitoring for hybrid pavement structure

Participants: Nicolas Le Touz, Thibaud Toullier, Jean Dumoulin.

In those studies, we evaluate by numerical modelling the energy inputs that could occur in a hybrid pavement structure with a semi-transparent or opaque wearing course bonded to a porous base layer, the seat of a heat transfer fluid circulation. The digital studies conducted propose a coupled resolution of various thermal phenomena: diffusion/convection in the case of an opaque surface drainage pavement, and diffusion/convection/radiation for a pavement with a semi-transparent surface. Coupled equation systems are solved numerically using the finite element method. This model was developed directly on a Matlab kernel. In a second time, laboratory experiments on small specimen were carried out and the surface temperature was monitored by infrared thermography. Results obtained are analyzed and performances of the numerical model for real scale outdoor application are discussed .[35], [34]

### 7.3.2. Phase change materials characterization

### Participant: Jean Dumoulin.

In a costs reduction and comfort requirements context, the use of phase change materials (PCM) is a sustainable and economical answer. For transportation infrastructures and winter maintenance, they avoid ice occurrence or snow accumulation. Their characteristics, and more specifically, the solid to liquid phase transition temperature and enthalpy, are usually obtained through DSC. Raman spectroscopy can bring answers and information on their microstructures. The liquid to solid phase change was investigated on three PCM, a paraffin, formic acid and diluted formic acid. A comparison made on freezing temperature obtained through DSC, Raman spectroscopy associated with chemiometrics indicated a consistency between the methods. Raman spectroscopy coupled with multivariate data analysis allowed the identification of an additional specificity in the freezing process of the paraffin. All methods provided results consistent between each other, although some differences between literature and experimental freezing temperatures of the considered PCM were observed in all cases. [20], [53]

### 7.4. Methods for building performance assessment

### 7.4.1. Building performance assessment

Participants: Jordan Brouns, Alexandre Nassiopoulos.

Two additive thermal sources are generally not simultaneously distinguishable from the only observation of their effect on the heat balance. However, there are cases where information about the variation regularity of these sources is known. This is typically the case of convective internal gains in the building, for which the use scenarios create discontinuous inputs while heat gains relating to the air leakage are regular in time. In the present paper, we introduce a method aiming to distinguish heat sources using this a priori knowledge about their dynamics. We provide numerical and experimental evidence that the method succeeds in separating/distinguishing these kind of sources. This method could be applied to the identification of the occupancy rate for measurement and verification plans or smart home systems such as learning thermostats. [16]

### 7.5. System identification

### 7.5.1. Variance estimation of modal parameters from subspace-based system identification

Participants: Michael Doehler, Laurent Mevel.

This work has been carried out in collaboration with Palle Andersen.

Subspace-based system identification allows the accurate estimation of the modal parameters (natural frequencies, damping ratios, mode shapes) from output-only measurements, amongst others with data-driven methods like the Unweighted Principal Component (UPC) algorithm. Due to unknown excitation, measurement noise and finite measurements, all modal parameter estimates are inherently afflicted by uncertainty. The information on their uncertainty is most relevant to assess the quality of the modal parameter estimates, or when comparing modal parameters from different datasets. A method for variance estimation is presented for the variance computation of modal parameters for the UPC subspace algorithm. Developing the sensitivities of the modal parameters with respect to the output covariances, the uncertainty is propagated from the measurements to the modal parameters from UPC. The resulting variance expressions are easy to evaluate and computationally tractable when using an efficient implementation. In a second step, the uncertainty information of the stabilization diagram is used to extract appropriately weighted global mode estimates and their variance. The method is applied to experimental data from the Z24 Bridge [30].

# 7.5.2. Bayesian parameter estimation for parameter varying systems using interacting Kalman filters

Participants: Antoine Crinière, Laurent Mevel, Jean Dumoulin, Subhamoy Sen.

This work is in collaboration with F. Cerou of ASPI team at Inria.

Standard filtering techniques for structural parameter estimation assume that the input force either is known exactly or can be replicated using a known white Gaussian model. Unfortunately for structures subjected to seismic excitation, the input time history is unknown and also no previously known representative model is available. A novel algorithm is proposed to estimate the force as additional state in parallel to the system parameters. Two concurrent filters are employed for parameters and force respectively, mixing interacting Particle Kalman filter and another filter employed to estimate the seismic force acting on the structure [38], [49].

### 7.5.3. From structurally independent local LTI models to LPV model

Participant: Qinghua Zhang.

This work on linear parameter varying (LPV) system identification has been carried out in collaboration with Lennart Ljung (Linköping University, Sweden).

The local approach to LPV system identification consists in interpolating individually estimated local linear time invariant (LTI) models corresponding to fixed values of the scheduling variable. It is shown in this work that, without any global structural assumption of the considered LPV system, individually estimated local state-space LTI models do not contain sufficient information for determining similarity transformations making them coherent. Nevertheless, it is possible to estimate these similarity transformations from input-output data under appropriate excitation conditions [21].

### 7.5.4. Stability of the Kalman filter for output error systems

Participant: Qinghua Zhang.

The stability of the Kalman filter is classically ensured by the uniform complete controllability regarding the process noise and the uniform complete observability of linear time varying systems. Recently we have studied the stability of the Kalman filter for output error (OE) systems, in which the process noise is totally absent. In this case the classical stability analysis assuming the controllability regarding the process noise is thus not applicable. Our first efforts were focused on continuous time systems, whereas discrete time systems have been studied since last year. It is shown in this work that the uniform complete observability is sufficient to ensure the stability of the Kalman filter applied to time varying OE systems, regardless of the stability of the OE systems [22].

### 7.5.5. Reduced-order interval-observer design for dynamic systems with time-invariant uncertainty

### Participant: Qinghua Zhang.

This work on interval-based state estimation has been carried out in collaboration with Vicenç Puig's team (Universitat Politècnica de Catalunya, Spain). The reported work addresses in particular the design of reduced-order interval-observers for dynamic systems with both time-invariant and time varying uncertainties. Because of the limitations of the set-based approach and the wrapping effect to deal with interval-observers, the trajectory-based interval-observer approach is used with an appropriate observer gain. Due to difficulties to satisfy the conditions for selecting a suitable gain to guarantee the positivity of the resulting observer, a reduced-order observer is designed to increase the degree of freedom when selecting the observer gain and to reduce the computational complexity. Simulation examples illustrates the effectiveness of the proposed approach [37].

# 7.5.6. Parameter uncertainties quantification for finite element based subspace fitting approaches

Participants: Guillaume Gautier, Laurent Mevel, Michael Doehler.

This work has been carried out in collaboration with Jean-Mathieu Mencik and Roger Serra (INSA Centre Val de Loire).

Recently, a subspace fitting approach has been proposed for vibration-based finite element model updating. The approach makes use of subspace-based system identification, where the extended observability matrix is estimated from vibration measurements. Finite element model updating is performed by correlating the model-based observability matrix with the estimated one. However, estimates from vibration measurements are inherently exposed to uncertainty. A covariance estimation procedure for the updated model parameters is proposed, which propagates the data-related covariance to the updated model parameters by considering a first-order sensitivity analysis. In particular, this propagation is performed through each iteration step of the updating minimization problem, by taking into account the covariance between the updated parameters and the data-related quantities. Simulated vibration signals and experimental data of a beam validate the method [18].

# 7.6. Damage diagnosis

# **7.6.1.** Damage detection by perturbation analysis and additive change detection theory **Participants:** Michael Doehler, Laurent Mevel, Oinghua Zhang.

The monitoring of mechanical systems aims at detecting damages at an early stage, in general by using output-only vibration measurements under ambient excitation. In this paper, a method is proposed for the detection and isolation of small changes in the physical parameters of a linear mechanical system. Based on a recent work where the multiplicative change detection problem is transformed to an additive one by means of perturbation analysis, changes in the eigenvalues and eigenvectors of the mechanical system are considered in the first step. In a second step, these changes are related to physical parameters of the mechanical system. Finally, another transformation further simplifies the detection and isolation problem into the framework of a linear regression subject to additive white Gaussian noises, leading to a numerically efficient solution of the considered problems. A numerical example of a simulated mechanical structure is reported for damage detection and localization [31].

### 7.6.2. Damage localization using the statistical subspace damage localization method

Participants: Michael Doehler, Laurent Mevel, Saeid Allahdadian.

This work is happening during a thesis in collaboration with C. Ventura at UBC, Vancouver.

In this paper the statistical subspace damage localization (SSDL) method is employed in localizing the damage in a real structure, namely the Yellow frame. The SSDL method is developed for real testing conditions and tested in two damage configurations. It was demonstrated that the SSDL method can localize the damage robustly in the Yellow frame for simple and multiple distinct damage scenarios using the analytical modal parameters. The method is described and its effectiveness is demonstrated [24].

# 7.6.3. Stochastic Subspace-Based Damage Detection with Uncertainty in the Reference Null Space

Participants: Michael Doehler, Laurent Mevel, Eva Viefhues.

This work is happening during a thesis in collaboration with F. Hille at BAM, Berlin.

This paper deals with uncertainty considerations in damage diagnosis using the stochastic subspace-based damage detection technique. With this method, a model is estimated from data in a (healthy) reference state and confronted to measurement data from the possibly damaged state in a hypothesis test. Previously, only the uncertainty related to the measurement data was considered in this test, whereas the uncertainty in the estimation of the reference model has not been considered. We derive a new test framework, which takes into account both the uncertainties in the estimation of the reference model as well as the uncertainties related to the measurement data. Perturbation theory is applied to obtain the relevant covariances. In a numerical study the effect of the new computation is shown, when the reference model is estimated with different accuracies, and the performance of the hypothesis tests is evaluated for small damages. Using the derived covariance scheme increases the probability of detection when the reference model estimate is subject to high uncertainty, leading to a more reliable test [41].

### 7.6.4. Statistical damage localization with stochastic load vectors

Participants: Md Delwar Hossain Bhuyan, Michael Doehler, Laurent Mevel, Guillaume Gautier.

This work is in collaboration with F. Schoefs and Y. Lecieux, GEM, Nantes.

The Stochastic Dynamic Damage Locating Vector (SDDLV) method is a damage localization method based on both a Finite Element (FE) model of the structure and modal parameters estimated from measurements in the damage and reference states of the system. A vector is obtained in the null space of the changes in the transfer matrix from both states and then applied as a load vector to the model. The damage location is related to this stress where it is close to zero. An important theoretical limitation was that the number of modes used in the computation could not be higher than the number of sensors located on the structure. In this paper, the SDDLV method has been extended with a joint statistical approach for multiple mode sets, overcoming this restriction on the number of modes. Another problem is that the performance of the method can change considerably depending of the Laplace variable where the transfer function is evaluated. Particular attention is given to this choice and how to optimize it. The new approach is validated in numerical simulations and on experimental data. From these results, it can be seen that the success rate of finding the correct damage localization is increased when using multiple mode sets instead of a single mode set [15], [52], [27].

### 7.6.5. Transfer matrices-based statistical damage localization and quantification

Participants: Md Delwar Hossain Bhuyan, Michael Doehler, Laurent Mevel, Guillaume Gautier.

This work is in collaboration with GEM, Nantes and C. Ventura at UBC, Nantes.

Vibration measurements and a finite element model are used to locate loss of stiffness in a steel frame structure at the University of British Columbia. The Stochastic Dynamic Damage Locating Vector (SDDLV) is compared to a sensitivity based approach developed by the authors. Both approaches have in common to be built on the estimated transfer matrix difference between reference and damaged states. Both methods are tested for localization and quantification on a structure at University of British Columbia [26], [28].

### 7.6.6. Statistical damage localization based on Mahalanobis distance

Participant: Michael Doehler.

This work is in collaboration with Aalborg University, Structural Vibration Solutions and Universal Foundation in Denmark during the thesis of S. Gres (Aalborg University).

In this paper, a new Mahalanobis distance-based damage detection method is studied and compared to the wellknown subspace-based damage detection algorithm. Methods are implemented using control charts to enhance the resolution of the damage detection. The damage indicators are evaluated based on the ambient vibration signals from numerical simulations on a novel offshore support structure and experimental example of a full scale bridge. The results reveal that the performance of the two damage detection methods is similar, hereby implying merit of the new Mahalanobis distance-based approach, as it is less computationally complex [32].

### 7.6.7. On the value of Information for SHM

Participant: Michael Doehler.

This work is issued from the COST Action TU1402.

The concept of value of information (VoI) enables quantification of the benefits provided by structural health monitoring (SHM) systems in principle. Its implementation is challenging, as it requires an explicit modelling of the structural system's life cycle, in particular of the decisions that are taken based on the SHM information. In this paper, we approach the VoI analysis through an influence diagram (ID), which supports the modelling process. We provide a simple example for illustration and discuss challenges associated with real-life implementation [39].

### 7.6.8. Structural system reliability and damage detection information

Participant: Michael Doehler.

This work is in collaboration with S. Thöns (DTU) during the thesis of L. Long (BAM).

This paper addresses the quantification of the value of damage detection system and algorithm information on the basis of Value of Information (VoI) analysis to enhance the benefit of damage detection information by providing the basis for its optimization before it is performed and implemented. The approach of the quantification the value of damage detection information builds upon the Bayesian decision theory facilitating the utilization of damage detection performance models, which describe the information and its precision on structural system level, facilitating actions to ensure the structural integrity and facilitating to describe the structural system performance and its functionality throughout the service life. The structural system performance is described with its functionality, its deterioration and its behavior under extreme loading. The structural system reliability given the damage detection information is determined utilizing Bayesian updating. The damage detection performance is described with the probability of indication for different component and system damage states taking into account type 1 and type 2 errors. The value of damage detection information is then calculated as the difference between the expected benefits and risks utilizing the damage detection information is determined, demonstrating Pratt truss system, the value of damage detection information is determined, demonstrating the potential of risk reduction and expected cost reduction [36].

# 7.6.9. Estimation of a cable resistance profile with readaptation of mismatched measurement instrument

Participants: Nassif Berrabah, Qinghua Zhang.

As the cumulative length of electric cables in modern systems is growing and as these systems age, it becomes of crucial importance to develop efficient tools to monitor the condition of wired connections. Therein, in contrast to hard faults (open or short circuits), the diagnosis of soft-faults requires a particular effort. Indeed, these faults are more difficult to detect, yet they are sometimes early warning signs of more important failures. In a previous paper, we proposed a method to compute the resistance profile of a cable from reflectometry measurements made at both ends of the cable. It enables detection, localization and estimation of dissipative soft-faults. In this reported work, we address the problem of impedance mismatch between the measurement instrument and the cable, based on a pre-processing of the measured data before running the estimation computations. It aims at reducing the impedance mismatch between instrumentation and the cable under test without physical intervention on the test fixtures. In addition, a measurement procedure has been developed in order to get the two-ends reflectometry measurements without actually connecting both ends of the cable under test to a single instrument [25].

## 7.7. Sensor and hardware based research

# 7.7.1. Cracks detection in pavement by a distributed fiber optic sensing technology

### Participant: Xavier Chapeleau.

This paper presents the feasibility of damage detection in asphalt pavements by embedded fiber optics as a new non-destructive inspection technique. The distributed fiber optic sensing technology based on the Rayleigh scattering was used in this study. The main advantage of this technique is that it allows to measure strains over a long length of fiber optic with a high spatial resolution, less than 1 cm. By comparing strain profiles measured at different times, an attempt was made to link strain changes with the appearance of damage (cracking) in the pavement. This non-destructive method was evaluated on accelerated pavement testing facility, in a bituminous pavement. In our experimentation, the optical fibers were placed near the bottom of the asphalt layer. The application of 728 000 heavy vehicle loads (65 kN dual wheel loads) was simulated in the experiment. Optical fiber measurements were made at regular intervals and surface cracking of the pavement was surveyed. After some traffic, a significant increase of strains was detected by the optical fibers at different points in the

pavement structure, before any damage was visible. Later, cracking developed in the zones where the strain profiles were modified, thus indicating a clear relationship between the increased strains and crack initiation. These first tests demonstrate that distributed fiber optic sensors based on Rayleigh scattering can be used to detect crack initiation and propagation in pavements, by monitoring strain profiles in the bituminous layers [17].

### 7.7.2. Wireless sensors and GPS synchronization

Participants: Vincent Le Cam, David Pallier.

Most of recent development in WSN domain focused on energy (saving or harvesting), on wireless protocols, on embedded algorithms. But it is a fact that, most of monitoring applications need samples to be timestamped. According to the application, the wished time resolution could be up to one second for automation monitoring, one millisecond for vibration, one microsecond for acoustic monitoring, one nanosecond for electricity or light propagation... The consequence for a Wireless network of electronic nodes is that, by nature, no common signal could physically provide a synchronization top. But, as each electronic device, a wireless sensor time-base uses a timer incremented by a quartz whose initial value is theoretical up to some p.p.m. and whose period drift on time because of age, temperature,... Two kind of solutions could be regarded : a synchronization signal provided by the wireless protocol itself; an absolute synchronization from a referential source such as: GPS, Frankfurt clock, Galileo,... In the first way, it will be demonstrated the poor accuracy and the need of energy such a mechanism offers. In the second way, the article will details how a deterministic (Universal Time), accurate and resilient algorithm has been implemented. The article also provides specific results of application on acoustic monitoring system and electricity propagation where the accuracy of a WSN has reached up to 10 nanosecond UT. Consequence on energy consumption of this algorithm are given with a description of future works to improve the energy balance while keeping the device sober and synchronized [33].

# 8. Bilateral Contracts and Grants with Industry

# 8.1. Bilateral Contracts with Industry

### 8.1.1. Contracts with SVS

Participants: Laurent Mevel, Michael Doehler.

I4S is doing technology transfer towards SVS to implement I4S technologies into ARTEMIS Extractor Pro. This is done under a royalty agreement between Inria and SVS.

In 2014, the damage detection toolbox has been launched http://www.svibs.com/products/ ARTeMIS\_Modal\_Features/Damage\_Detection.aspx.

In 2015, SVS and Inria have earned an Innobooster grant to help transfer algorithms in 2016 Artemis Extractor Pro.

In 2016, uncertainty quantification for modal analysis has been launched.

In 2017, a new Innobooster grant has been obtained for the uncertainty analysis of mode shapes in Artemis.

### 8.1.2. Contract with SNCF: DEMETER

Participants: Vincent Le Cam, Quentin Bossard, Mathieu Le Pen.

IFSTTAR's engineers Arthur Bouche and Laurent Lemarchand are contributing to this project.

DEMETER is one of the major projects for I4S in terms of strategy, scientific and technological impact.

DEMETER is a meta project whose global objective is the validation of the contribution of the Internet of Things (IOT) applied to the health monitoring of railway items. SNCF and IFSTTAR have signed a roadmap for safety relevant items, where wireless monitoring and smart algorithms could bring strong improvements to SNCF in terms of real-time maintenance or predictive maintenance. Those items are, amongst others:

- Crossing engine motor monitoring
- Needle motor monitoring
- Axel counter monitoring
- Train detection pedal monitoring

In each case, a prototype of a specific wireless and smart sensor is designed (that may or may not use PEGASE 2 platform), installed along railway lines in service and data are transmitted wirelessly to the cloud supervisor at IFSTTAR for evaluation in SHM algorithms.

In particular, during 2017 SNCF and IFSTTAR have performed following common works:

- finalization of the TRAIN PEDAL DETECTION instrumentation with smart sensors using new wireless and industrial IOT protocoles: LoRa and Sigfox. A specific pedal is now subject of in situ test led by SNCF
- axel counter monitoring has been the major R&D subject of 2017: 2 entire and specific smart sensors have been designed, programmed and installed at Chevilly specific SNCF testbench (e.g. with real train passages). Specific algorithms (such as PID and Pattern Recognition) have been modeled and programmed into PEGASE2 platform for these new sensors.

For the future, new projects related to

- water-level monitoring around railways has been setup
- ballast vibration monitoring of railways has been setup
- "unshunting of electrical lines at train passage" detection around railways

have been initiated with SNCF R&D department.

### 8.1.3. Contracts with SDEL-CC (VINCI Group)

Participants: Vincent Le Cam, Mathieu Le Pen.

This work was done in collaboration with Laurent Lemarchand, and Arthur Bouche at IFSTTAR, SII, Nantes.

Following a 2016 contract, a new contract was signed in 2017 until end 2018, with the company SDEL-CC, a 100% daughter of the VINCI Group, Energy department. The project exploits the unique time stamp capacity of the PEGASE 2 platform up to 50 nanoseconds, independently of distances in the network of PEGASE2 nodes. The synchronization capacity is employed to design a sensor prototype based on PEGASE 2 to time-stamp the current wave after a lightning impact on a high-voltage line. By knowing the exact time, the wave can be seen at each extremity of the electrical line to localize accurately the lightning impact point.

During 2017, a real high-voltage electrical line has been instrumented: at each end of the line, 2 sensors have been set up and data are sent in real time to a cloud platform. Furthermore, the software of the platform was optimized: at the embedded level (i.e. on PEGASE 2 wireless system) with new algorithms to correct time synchronization up to some 10 nanoseconds, at the cloud level with a specific QT C++ Interface to display results (i.e. lightning localization on electrical line) and to transform raw data into ComTrade standard representation

Discussions are ongoing with SDEL-CC to transform the prototype into a future product. In 2017 it has to be mentioned that the project has been submitted to VINCI International challenges (over 150 000 collaborators) and has been awarded with The Best Vinci Innovation Award.

# 8.2. Bilateral Grants with Industry

### 8.2.1. PhD project with EDF – Electrical device ageing monitoring

Participants: Nassif Berrabah, Qinghua Zhang.

A joint PhD project between Inria and EDF (Electricité de France) was started in December 2014 and finished in November 2017 with Nassif Berrabah's PhD thesis defense. The purpose of this study is to develop methods for the monitoring of electrical instruments in power stations, in order to prevent failures caused by ageing or accidental events. This project has been funded by EDF and by the ANRT agency. The main outcome of this project is an efficient reflectometry-based method for resistive fault detection, localization and quantification, capable of dealing with both distributed and localized faults, with associated data processing tools taking into account practical constraints in industrial applications. These results have led to a patent jointly filed by EDF and Inria.

# 9. Partnerships and Cooperations

## 9.1. Regional Initiatives

### 9.1.1. MONEOL

Participants: Ivan Guéguen, Guillaume Gautier, Laurent Mevel.

Type: CEAtech PDL

Objectif: Modal analysis of wind turbines using new sensors

Duration: 09/2015 to 11/2017.

Coordinator: Louis Marie Cotineau (IFSTTAR)

Inria contact: Guillaume Gautier

Abstract: The MONEOL project aims to demonstrate the feasibility of using Morphosense as a vibration monitoring system for wind turbines. It is proposed to set up a demonstrator consisting of a monitoring system placed in the mast of the wind turbine, a vibration analysis system and a visualization of the vibratory state at the CEA-Tech premises, located on the Technocampus Ocean of Nantes allowing to visualize in real time (quasi) the modal deformations of the mast of the wind turbine.

The MONEOL (Wind turbine monitoring) project was concluded in September 2017. Morphosense ribbon was deployed on the mast of a wind turbine. Morphosense validation was conducted through the comparison with a classical vibration monitoring system. The relevance regarding the use of such a system was highlighted, especially due to reduced installation time. SSI algorithms, including modal parameters and damage identification, were implemented inside the Morphosense. Actually, wind turbine health condition is displayed in real time through a web page.

# 9.1.2. Interactive Communication (InterCom): Massive random access to subsets of compressed correlated data

Participants: Jean Dumoulin, Antoine Crinière, Frederic Gillot.

Type: Labex COMINLABS

Objectif: Massive random access to large-scale sensor network (Smart Cities)

Duration: Since November 2016 to Nov. 2019.

Coordinator : Aline Roumy, Thomas Maugey (Sirocco), Jean Dumoulin (I4S)

Partners: Elsa Dupraz (Lab-STICC), Aline Roumy (IRISA, Sirocco team), Michel Kieffer (L2S), Thomas Maugey(IRISA, Sirocco team), CentraleSupelec, Univ. Paris Sud.

### Inria contact: Jean Dumoulin

Abstract: This project aims to develop novel compression techniques allowing massive random access to large databases. Indeed, we consider a database that is so large that, to be stored on a single server, the data have to be compressed efficiently, meaning that the redundancy/correlation between the data have to be exploited. The dataset is then stored on a server and made available to users that may want to access only a subset of the data. Such a request for a subset of the data is indeed random, since the choice of the subset is user-dependent. Finally, massive requests are made, meaning that, upon request, the server can only perform low complexity operations (such as bit extraction but no decompression).

Algorithms for two emerging applications of this problem will be developed: Free-viewpoint Television (FTV) and massive requests to a database collecting data from a large-scale sensor network (such as Smart Cities) in which I4S is involved.

### 9.1.3. MAG2C-Pont Tabarly

Participants: Ivan Guéguen, Jean Dumoulin.

Type: GIS Objectif: bridge instrumentation Duration: Since 2014 Coordinator: LIRGEC Partners: IFSTTAR, CSTB, Nantes Métropole, Université de Nantes Inria contact: Ivan Guéguen

Abstract: The project deals with the instrumentation of the Tabarly Bridge. In collaboration with Nantes Métropole, CSTB, and Université de Nantes, instrumentation of both dynamical and InfraRed properties of an operational bridge are investigated. These measures coupled with a wireless data transmission system will allow remote monitoring of the evolution of the structure. Objective is to couple different kind of measurement to achieve thermo-vibration monitoring of the structure. This is a big milestone for the team and our objective to mix thermo-vibration data.

# 9.1.4. MAG2C-MOSIWIND (MOnitoring of Structural Integrity of an onshore WIND turbine slab foundation and tower)

Participants: Xavier Chapeleau, Ivan Guéguen.

Type: GIS

Objectif: MOnitoring of Structural Integrity of an onshore WIND turbine slab foundation and tower Duration: Since 2015

Coordinator : LIRGEC

Partners: IFSTTAR, CSTB, Nantes Métropole, Université de Nantes, ECN, Valorem, Valréa and Valémo

#### Inria contact: Xavier Chapeleau

Abstract: The project deals with the instrumentation of an onshore WIND turbineâs slab foundation and tower. The aim is to experiment sensors and methods for structural integrity monitoring of an onshore wind turbine under real conditions and to qualify them over long term. Before casting, the concrete slab foundation (20m in diameter, 3.85m high, 450m3 of concrete, 48T of reinforcement) was first instrumented with continuous optical fibers, optical strain gauges, temperature sensors and accelerometers. Afterwards, accelerometers were placed in the mast. Data obtained by these different sensors will help, on the one hand, to monitor changes in the dynamic behavior of the structure in order to verify that they remain within the limits fixed during the design and, on the other hand, to detect any damage that could be critical for the safety of the structure. For this, SSI methods under ambient vibration will be applied. Since July 2017, only the data of accelerometers measurements are logged periodically. The installation of systems of measurements for distributed fiber optics sensors and optical strains gauges remains to be done as soon as it can be possible to access to the wind turbine.

### 9.1.5. Collaboration with GeM

Participants: Laurent Mevel, Michael Doehler, Md Delwar Hossain Bhuyan.

Md Delwar Hossain Bhuyan has done his PhD on damage localisation on civil structures in collaboration with GeM (Institute of Civil and Mechanical Engineering), Université de Nantes. The thesis is co-directed by L. Mevel, and F. Schoefs from GeM, with supervision shared with M. Doehler and Y. Lecieux from GeM. It is funded by the Brittany region for 3 years and has been successfully defended in November 2017.

### 9.1.6. Collaboration with IETR

Participants: Vincent Le Cam, David Pallier.

The thesis is directed by Sébastien Pillement at IETR. It is funded by RFI WISE Electronique Professionnelle within the SENTAUR project.

The subject of the thesis is to study, implement and propose a deterministic and reliable dating solution for wireless sensor networks. This solution must take into account both the risks of loss of synchronization signals, environmental hazards and the desire to achieve the most sober possible solution in energy.

## **9.2.** National Initiatives

### 9.2.1. High speed rail track instrumentation

Participant: Ivan Guéguen.

#### Type: IRT

Objective: rail track SHM

Duration: 11/2014 to 11/2018

Coordinator: RAILENIUM

Partners: IFSTTAR, EIFFAGE, RFF, LGCgE

Inria contact: Ivan Guéguen

Abstract: This project aims to orchestrate multiple sections of a high-speed route (classical section with granular layer, transition zone). The proposed instrumentation concerns all the different layers of the structure, and is designed to allow monitoring of the overall track behavior. Using accelerometers and weather station, this instrumentation will estimate the fatigue life and temperature changes in the track.

## 9.2.2. ANR Resbati

Participants: Ludovic Gaverina, Jean Dumoulin.

### Type: ANR

Objectif: In-situ measurements of thermal wall resistance

Duration: 10/2016 to 10/2019

Coordinator: Laurent Ibos

Partners : IFSTTAR, CERTES, CEREMA, CSTB, LNE, THEMACS, AFNOR

Inria contact: Jean Dumoulin

Abstract: RESBATI is an applied research project whose objective is to develop a field measurement device that meets precise specifications to systematically measure the level of thermal insulation of building walls. The preferred metrological tool is infrared thermography.

### 9.2.3. Equipex Sense-City

Participants: Jean Dumoulin, Laurent Mevel, Antoine Crinière.

Through the ADT Cloud2SM, participation of I4S in SenseCity was possible. IFSTTAR's SensorBox developed by Jean Dumoulin was installed and presented at SEnseCity Kick off and is installed on-site. Cloud2IR and Cloud2SM software have been deployed within the ADT of A. Crinière. (http://sense-city.ifsttar.fr/)

# 9.3. European Initiatives

### 9.3.1. FP7 & H2020 Projects

9.3.1.1. Built to Specifications (Built2Spec)

Participant: Jean Dumoulin.

Type: Horizon 2020

Defi: Model Driven Physical Systems Operation

Objectif: Reduce the gap between a building?s designed and as-built energy performance.

Duration: January 2015 to January 2019

Coordinator: Manager and project head : NOBATEK, Germain Adell. For CERMA : Marjorie Musy Inria teams I4S

Inria contact: J. Dumoulin

Partners: Consortium of 20 Public and Industrial actors

Website: http://built2spec-project.eu/

Abstract: Built to Specifications (Built2Spec) is involving 20 European partners that seeks to reduce the gap between a building designed and as-built energy performance. To do this, the project put a new set of breakthrough technological advances for self-inspection checks and quality assurance measures into the hands of construction professionals. The project aims to deliver Building Information Modelling (BIM) and Thermal and 3D Imaging Tools among others.

The project is in collaboration with formers members of the team, Alexandre Nassiopoulos and Jordan Brouns, now working at Ecotropy, SME.

9.3.1.2. INFRASTAR(Innovation and Networking for Fatigue and Reliability Analysis of Structures – Training for Assessment of Risk)

Participants: Xavier Chapeleau, Antoine Bassil.

Call: H2020-MSCA-ITN-2015 (Horizon 2020 â Marie-Sklodowska Curie Actions â Innovative Training Networks)

Type of Action: MSCA-ITN-ETN

Objectif: Reduce the gap between a building?s designed and as-built energy performance.

Duration: 48 months since 2016 May 1st

Coordinator: Odile Abraham (IFSTTAR)

Academic and industrial Partners: IFSTTAR, UNIVERSITY OF AALBORG, BAM, EPFL, GuD Consult Gmbh, COWI A/S, NeoStrain, PHIMECA

Inria contact: X. Chapeleau

Website: http://infrastar.eu/

Abstract: The aim of INFRASTAR project is to develop tools combining modeling and measurements for the prediction of the fatigue behavior of concrete structures (bridges and foundations of wind turbines) with the ultimate objective of establishing an efficient strategy for inspection and reinforcement operations. In the second half of 2016, 12 young researchers were recruited to carry out and cross-examine research on monitoring and auscultation (WP 1), structural models (WP 2) and reliability of approaches for decision-making (WP 3). In this project, a phd student (Antoine Bassil) was recruited (Nov. 2016) on the fatigue monitoring of concrete structure by fibre-optic sensors. During the first 6 months of the thesis, a State of the Art on the use of fiber optic sensors for structural health monitoring in civil engineering was done and mostly by focusing on distributed optical fiber sensor's technology (DOFS) for crack detection in concrete. This State of the Art shows that distributed optical fiber sensor can localize accurately cracks in concrete if they propagate across the sensor. However, the quantification of the crack widths by distributed optical fiber sensor remains a scientific challenge. Indeed, it is necessary to take into account of the mechanical strain transfer of the fiber sensor. Now, the second part of the phd student work is to develop a theoretical model for the mechanical strain transfer function and to validate it by experimental tests. The main milestone of the modelling to overcome is to take into account of slippage and elasto-plastic effects

### 9.3.2. Collaborations in European Programs, Except FP7 & H2020

### 9.3.2.1. COST Action TU 1402

Participants: Michael Doehler, Laurent Mevel.

L. Mevel is member of the management committee of the COST Action.

M. Doehler is co-leader of working group 2 "SHM strategies and structural performance" and member of the steering committee.

### Type: COST

Objectif: Quantifying the value of structural health monitoring

Duration: 11/2014 - 11/2018

Coordinator: S. Thoens (DTU Denmark)

Partner: 23 countries, see http://www.cost.eu/COST\_Actions/tud/Actions/TU1402

Inria contact: Laurent Mevel

Abstract: Since 2014, until 2018, the COST Action has altogether around 120 participants from over 25 countries. This Action aims to develop and describe a theoretical framework, together with methods, tools, guidelines, examples and educational activities, for the quantification of the value of SHM. Progress of the action is presented in [40].

#### 9.3.2.2. PROCOPE 37826QE

Participants: Michael Doehler, Laurent Mevel, Eva Viefhues.

#### Type: PHC PROCOPE

Objectif: Statistical damage localization for civil structures

Duration: 01/2017 - 12/2018

Coordinator: M. Doehler

Partner: BAM German Federal Institute for Materials Research and Testing

Inria contact: M. Doehler

Abstract: Our main objective is the development of a theoretically solid damage localization method that does not only work in simulations and lab experiments, but on structures in the field under real operational conditions. This German-French mobility grant is in support of Eva Viefhues' PhD thesis.

### 9.3.3. Collaborations with Major European Organizations

# 9.3.3.1. European Research Network on System Identification (ERNSI)

Participants: Qinghua Zhang, Michael Doehler, Laurent Mevel.

The I4S project-team is involved in the activities of the European Research Network on System Identification (ERNSI) federating major European research teams on system identification. Modeling of dynamical systems is fundamental in almost all disciplines of science and engineering, ranging from life science to process control. System identification concerns the construction, estimation and validation of mathematical models of dynamical physical or engineering phenomena from experimental data.

### 9.3.4. Other European Programs

### 9.3.4.1. Innobooster

Participants: Michael Doehler, Laurent Mevel.

Together with SVS, we got the Danish Innobooster innovation grant for industrial research and transfer. In 2017, the grant was awarded to transfer methods for the identification of mode shapes and their uncertainty [30] to SVS' ARTeMIS software.

# 9.4. International Initiatives

### 9.4.1. Informal International Partners

9.4.1.1. Collaboration with CNR, Italy

Participants: Jean Dumoulin, Nicolas Le Touz.

Non destructive testing on outdoor structures by coupling infrared thermography with ground penetrating radar is one of the topic addressed in this collaboration. A new one about TerHertz is starting. A proposal for associated lab is currently drafted.

9.4.1.2. Collaboration with British Columbia University, Canada

Participants: Laurent Mevel, Michael Doehler, Saeid Allahdadian.

Saeid Allahdadian was PhD student of professor Carlos Ventura in Vancouver. Following our recent papers, Michael Doehler has been invited to co-supervise the PhD of Saeid Allahdadian. The thesis has been defended this year.

9.4.1.3. Collaboration with BAM, Germany

Participants: Laurent Mevel, Michael Doehler, Eva Viefhues.

Eva Viefhues is currently PhD student of Laurent Mevel and Michel Doehler in Berlin, financed by BAM. M. Doehler is also associate researcher of the BAM institut since 2016.

9.4.1.4. Collaboration with Politecnico di Milano, Italy Participants: Michael Doehler, Francesco Giordano.

During COST Action TU 1402 and professor M.P. Limongelli's research stay at IFSTTAR, collaboration with Politecnico di Milano has started, resulting in several joint publications in 2016 and 2017. PhD student F. Giordano has started at Politecnico Milano in November 2017 under the direction of M.P. Limongelli, M. Doehler is co-supervising.

9.4.1.5. Collaboration with Technical University of Denmark (DTU) Participant: Michael Doehler.

During COST Action TU 1402 and previously at BAM, collaboration with Sebastian Thöns from DTU in Denmark started on risk analysis and SHM based reliability updating. Also, DTU's PhD student Lijia Long is involved [36].

9.4.1.6. Collaboration with Aalborg University, Denmark **Participant:** Michael Doehler.

Together with Structural Vibration Solutions, collaboration with Aalborg University (professor Lars Damkilde, Department of Civil Engineering) has started during the PhD of Szymon Gres on damage detection methods [32].

## 9.5. International Research Visitors

### 9.5.1. Visits of International Scientists

Within the PROCOPE mobility grant, E. Viefhues and F. Hille from BAM, Germany, visited Inria for one week each.

Prof. Xingwen Liu from Southwest University of China visited the I4S team for one week.

### 9.5.2. Visits to International Teams

Within the PROCOPE mobility grant and the collaboration in L. Long's PhD thesis with S. Thöns, M. Doehler spent 4 weeks at BAM, Germany.

Within the IFSTTAR foreign affair department grant and the existing collaboration, J. Dumoulin spent 2 weeks at Laval University, Canada.

Within the H2020 INFRASTAR Project A. Bassil spent 3 months at BAM.

# **10.** Dissemination

# **10.1. Promoting Scientific Activities**

### 10.1.1. Scientific Events Selection

10.1.1.1. Member of the Conference Program Committees

J.Dumoulin is

- member of the scientific committee of the GI Division (Geosciences Instrumentation and Data Systems) of EGU (European Geosciences Union) for infrastructure instrumentation and monitoring since April 2013. (http://www.egu.eu/gi/structure/)
- member of the scientific committee of QIRT (quantitative Infrared Thermography) since February 2014 (http://www.qirt.org/)
- organizer and chair of a session at EGU 2017 (http://www.egu2017.eu/).

Q. Zhang is

- member of the international program committee of the 18th IFAC Symposium SYSID that will take place in Stockholm, Sweden, July 9-11, 2018.
- member of the international program committee of the 10th IFAC Symposium SAFEPROCESS that will take place in Warsaw, Poland, 29-31 August 2018.
- member of IFAC Technical Committee on Modelling, Identification and Signal Processing.
- member of IFAC Technical Committee on Fault Detection, Supervision and Safety of Technical Processes.
- member of IFAC Technical Committee on Adaptive and Learning Systems.

### L. Mevel is

- member of the EWSHM scientific committee.
- member of the IOMAC scientific committee.
- V. Le Cam is
  - member of the IWSHM scientific committee.
  - head and general secretary of the EWSHM scientific committee.
- M. Doehler is member of IFAC Technical Committee on Modelling, Identification, and Signal Processing.

### 10.1.1.2. Reviewer

V. Le Cam was session chairman for IWSHM 2017 in Stanford

L. Mevel was session chairman for IWSHM 2017 in Stanford

Q. Zhang was reviewer for CDC 2017, ACC 2018.

M. Doehler was session organizer at a COST workshop (http://www.cost-tu1402.eu/), session chairman at IOMAC 2017, and reviewer for ACC 2018.

J. Dumoulin was reviewer for QIRT ASIA 2017 and session chairman at EGU 2017 in GI division

### 10.1.2. Journal

### 10.1.2.1. Member of the Editorial Boards

L. Mevel is member of the editorial board of the journal Mathematical Problems in Engineering, and of the journal Shock and Vibration.

Q. Zhang is member of the editorial board of the journal of Intelligent Industrial Systems.

J. Dumoulin is member of the editorial board of the journal Quantitative Infrared Thermography, and of the journal Geoscientific Instrumentation and Data Systems.

### 10.1.2.2. Reviewer - Reviewing Activities

L. Mevel was reviewer for Mechanical Systems and Signal Processing, journal of Sound And Vibration, Sensors, Advanced Engineering Informatics, Structural Control and Health Monitoring, Advances in Structural Engineering

M. Doehler was reviewer for Automatica, Mechanical Systems and Signal Processing, Journal of Sound and Vibration, Journal of Testing and Evaluation

J. Dumoulin was reviewer for Quantitative Infrared Thermography Journal, GI Journal (EGU), SFT conference, ASME New NDE Journal

F. Gillot was reviewer for Structural and Multidisciplinary Optimization, Applied Mathematical Modelling, Shock and Vibration, Applied Sciences

### 10.1.3. Invited Talks

M. Doehler and L. Mevel, "Méthodes statistiques pour l'analyse vibratoire des structures," Journée scientifique Évaluation non destructive dans le génie civil de l'énergie, Nantes, France

M. Doehler, "Subspace-based methods for damage assessment," Wölfel Engineering, Würzburg, Germany

J. Dumoulin, "Infrared thermography in civil engineering: from non destructive testing in laboratory to outdoor thermal monitoring", QIRT ASIA 2017, Daejeon, Soth Korea

N. Le Touz, J. Dumoulin and J-M. Piau, "Etude et développement d'un modèle EF de transfert de chaleur multi-physique : application à l'étude de routes solaires hybrides", Journée thématique du groupe rayonnement de la société Française de Thermique sur "Méthodes numériques pour la résolution de l'équation de transfert radiatif : développements récents, modèles, et objectifs", Paris, France

V. Le Cam, "Internet of Things and new chalenges for Transportation and Structural Monitoring", plenary talk given to Committee of Transport Minestry, on 25 january 2017, to CEA List, on 16th March 2017, and to assembly of COFREND (500 people) on 30th May 2017

### 10.1.4. Research Administration

V. Le Cam is member of the scientific council of WEN (West Electronic Network) since 2014, which is a cluster of about 200 companies, academics and research laboratories active in electronics. During 2017, he has been involved amongst others in meetings and selection of R&D projects, PhD and post-doc funding, international mobility.

M. Doehler was reviewer of a research proposal submitted to the Polish National Science Centre (national research agency).

## **10.2. Teaching - Supervision - Juries**

### 10.2.1. Teaching

- J. Dumoulin
  - Licence Professionnelle TAM : thermographie infrarouge active, 16h, Université Paris-Est, France
  - Master 2 MMMRI (Maintenance et Maîtrise des Risques Industriels), contrôle non destructif par thermographie infrarouge active, 12h, Université Paris-Est, France
  - Master 2 ITII, BTP, module Maintenance et réhabilitation des ouvrages, Transferts thermiques dans les Structures : Des principes physiques à l'application sur site réel, 12 h, Ecole Centrale de Nantes(ECN), France.
  - Conference course, 2h, IR inspection of infrastructures: Scope of application, technical solutions and analysis methods, QIRT ASIA 2017, Daejeon, South Korea
V. Le Cam

- Master 2 Civil engineering, Structural Monitoring, 4h, Université de Nantes, France
- Licence 3 Professional SEICOM, 3h of theoretical lessons and 20H of practical lessons on Embedded and Smart Systems, Université de Nantes, France
- ESEO, 16h, practical lessons on embedded and smart systems under Linux, France
- Master 2 Electrical Engineering (GEII), 4h on electronic systems and Structural Monitoring, Université Bretagne Sud, France

#### M. Doehler

- Master 1 informatique, 24 TD projet recherche, Université de Rennes 1 & ENS Rennes, France
- Cycle préparatoire intégré, STPI, mathématiques, 96h TD, INSA Rennes, France
- Conference course, 2h, Advanced Operational Modal Analysis using Stochastic Subspace Identification and its Applications, IOMAC 2017, Ingolstadt, Germany

X. Chapeleau

• Licence Pro Mesures physiques, Mesures optiques, 15h, IUT de St Nazaire, Université de Nantes, France

T. Toullier

- Master 1, TP Capteurs (12h), contrôle, commandes, École Centrale de Nantes, France
- Foundation Master, TD Programming and Data Analysis (14h), École Centrale de Nantes, France

F. Gillot

- Master 1, Conception optimale robuste de systèmes mécaniques (10h), École Centrale de Lyon, France
- Master 1, Dynamique des systèmes biologiques humains (4h), École Centrale de Lyon, France
- Formation initiale des ingénieurs de l'École Centrale de Lyon, TP, TD, BE, niveau L3, (50h), France

#### 10.2.2. Supervision

PhD : Antoine Bassil, *Fibre-optic sensor for fatigue monitoring*, D. Leduc, O. Abraham and X. Chapeleau. Ecole doctorale SPIGA, Université de Nantes, since November 2016.

PhD : Delwar Hossain Bhuyan, *Damage localisation for civil structures*, L. Mevel, F. Schoefs, Y. Lecieux and M. Doehler. Ecole doctorale MathSTIC, Université de Rennes 1, defended in 2017.

Guillaume Gautier's post-doctoral project on morphosense sensors, L. Mevel, 2015-2017.

Shubamoy Sen's post-doctoral project on seismic event monitoring, L. Mevel, 2016-2017.

Guillaume Gautier's post-doctoral project on seismic event monitoring, L. Mevel, 2017-2018.

Ludovic Gavérina post-doctoral project on in-situ measurement of thermal resistance of building envelopes, J. Dumoulin, march 2017- february 2019.

PhD : Nassif Berrabah, Electrical cable ageing monitoring , Q. Zhang, Ecole doctorale MathSTIC, Université de Rennes 1, defended in 2017.

PhD: Nicolas Le Touz. *Design and study of positive energy transport infrastructures: from thermomechanical modeling to the optimization of such energy systems* J. Dumoulin. Ecole Centrale Nantes (ECN) since december 2015.

PhD : Thibault Toullier. *Simultaneous characterization of the radiative properties and temperatures of envelopes of structures in natural environment by multispectral infrared thermography* L. Mevel, J. Dumoulin and M. Doehler. Ecole doctorale MathSTIC, Université de Rennes 1, since November 2016.

PhD : Saeid Allahdadian, *Methods for vibration-based damage assessment*, M. Doehler and C. Ventura. University of British Columbia, Canada, defended in 2017.

PhD : Eva Viefhues, *Statistical damage localization for civil structures*, L. Mevel and M. Doehler. Ecole doctorale MathSTIC, Université de Rennes 1, since November 2016.

PhD : Francesco Giordano, *Value of information for strain monitoring of infrastructure*, M. Doehler and MP. Limongelli and F. Bourquin. Politecnico Milano, Italy, since November 2017.

PhD : David Pallier, *Sensor Enhancement to Augmented Usage and Reliability*, S. Pillement, IETR, V. Le Cam, Ecole doctorale MathSTIC

J. Dumoulin is associate professor at Laval University, Canada.

M. Doehler is associate researcher at BAM, Germany.

#### 10.2.3. Juries

Jean Dumoulin was invited jury member for the PhD defense of Yingying YANG at I2M in Bordeaux.

### **10.3.** Popularization

The Hybrid solar road Mock-up (presented at the French Pavillon during COP21) has been invited and presented at:

- Forum National des Travaux Publics, Carroussel du Louvre (Paris), February 2017
- Innovation day des Travaux Publics, Casino du Lyon Vert (Lyon), December 2017
- Fête de la science, Ecole d'architecture de Nantes, October 2017

# 11. Bibliography

## Major publications by the team in recent years

- J. BROUNS, A. CRINIÈRE, J. DUMOULIN, A. NASSIOPOULOS, F. BOURQUIN. Diagnostic de structures de Génie Civil : Identification des propriétés spatiales et de la surface d'un défaut, in "SFT 2014", Lyon, France, Société Française de Thermique, May 2014, https://hal.inria.fr/hal-01082184.
- [2] A. CRINIÈRE, J. DUMOULIN, C. IBARRA-CASTANEDO, X. MALDAGUE. Inverse model for defect characterisation of externally glued CFRP on reinforced concrete structures: comparative study of square pulsed and pulsed thermography, in "Quantitative InfraRed Thermography Journal", March 2014, vol. 11, n<sup>o</sup> 1, p. 84-114 [DOI: 10.1080/17686733.2014.897512], https://hal.archives-ouvertes.fr/hal-01081174.
- [3] J. DUMOULIN, V. BOUCHER. Infrared thermography system for transport infrastructures survey with inline local atmospheric parameter measurements and offline model for radiation attenuation evaluations, in "Journal of Applied Remote Sensing", 2014, vol. 8, n<sup>o</sup> 1, p. 084978–084978.
- [4] J. DUMOULIN, A. CRINIÈRE, R. AVERTY. The detection and thermal characterization of the inner structure of the 'Musmeci' bridge deck by infrared thermography monitoring, in "Journal of Geophysics and Engineering", December 2013, vol. 10, n<sup>o</sup> 6, 17 [DOI: 10.1088/1742-2132/10/6/064003], https://hal.inria.fr/ hal-01081320.
- [5] M. DÖHLER, L. MEVEL. Fast Multi-Order Computation of System Matrices in Subspace-Based System Identification, in "Control Engineering Practice", September 2012, vol. 20, n<sup>o</sup> 9, p. 882–894.
- [6] M. DÖHLER, L. MEVEL. Modular Subspace-Based System Identification from Multi-Setup Measurements, in "IEEE Transactions on Automatic Control", November 2012, vol. 57, n<sup>O</sup> 11, p. 2951–2956.

- [7] M. DÖHLER, L. MEVEL. Efficient Multi-Order Uncertainty Computation for Stochastic Subspace Identification, in "Mechanical Systems and Signal Processing", June 2013, vol. 38, n<sup>o</sup> 2, p. 346–366.
- [8] M. DÖHLER, L. MEVEL. Subspace-based fault detection robust to changes in the noise covariances, in "Automatica", September 2013, vol. 49, n<sup>o</sup> 9, p. 2734–2743 [DOI: 10.1016/J.AUTOMATICA.2013.06.019], https://hal.inria.fr/hal-00907662.
- [9] A. JHINAOUI, L. MEVEL, J. MORLIER. A new SSI algorithm for LPTV systems: application to a hinged-bladed helicopter, in "Mechanical Systems and Signal Processing", January 2014, vol. 42, n<sup>o</sup> 1, p. 152–166.
- [10] P. LAIR, J. DUMOULIN, P. MILLAN. *Inverse method for flux characterization using infrared thermography in die forging*, in "Numerical Heat Transfer, Part A Applications", 1998, vol. 33, n<sup>o</sup> 3, p. 267–277.
- [11] F. LOETE, Q. ZHANG, M. SORINE. Experimental validation of the inverse scattering method for distributed characteristic impedance estimation, in "IEEE Transactions on Antennas and Propagation", 2015, vol. 63, n<sup>o</sup> 6, 7 [DOI: 10.1109/TAP.2015.2417215], https://hal.inria.fr/hal-01231807.
- [12] L. MARIN, M. DÖHLER, D. BERNAL, L. MEVEL. Robust statistical damage localization with stochastic load vectors, in "Structural Control and Health Monitoring", March 2015, vol. 22, n<sup>o</sup> 3.
- [13] M. ZGHAL, L. MEVEL, P. DEL MORAL.*Modal parameter estimation using interacting Kalman filter*, in "Mechanical Systems and Signal Processing", August 2014, vol. 47, n<sup>o</sup> 1, p. 139–150.
- [14] Q. ZHANG, M. SORINE, M. ADMANE. Inverse Scattering for Soft Fault Diagnosis in Electric Transmission Lines, in "IEEE Transactions on Antennas and Propagation", 2011, vol. 59, n<sup>o</sup> 1, p. 141 - 148, https://hal. inria.fr/inria-00365991.

### **Publications of the year**

#### **Articles in International Peer-Reviewed Journal**

- [15] M. D. H. BHUYAN, M. DÖHLER, Y. LECIEUX, L. MEVEL, F. SCHOEFS. Statistical damage localization with stochastic load vectors using multiple mode sets, in "Structural Health Monitoring", August 2017, vol. 16, n<sup>o</sup> 5, p. 518-535 [DOI: 10.1177/1475921717714447], https://hal.inria.fr/hal-01577250.
- [16] J. BROUNS, A. NASSIOPOULOS, K. LIMAM, F. BOURQUIN. Heat source discrimination in buildings to reconstruct internal gains from temperature measurements, in "Energy and Buildings", January 2017, vol. 135, p. 253-262 [DOI: 10.1016/J.ENBUILD.2016.11.041], https://hal.archives-ouvertes.fr/hal-01409762.
- [17] X. CHAPELEAU, J. BLANC, P. HORNYCH, J.-L. J. GAUTIER, J. CARROGET. Assessment of cracks detection in pavement by a distributed fiber optic sensing technology, in "Journal of Civil Structural Health Monitoring", 2017, vol. 16 [DOI: 10.1016/J.PSCR.2014.08.002], https://hal.inria.fr/hal-01582525.
- [18] G. GAUTIER, L. MEVEL, J.-M. MENCIK, R. SERRA, M. DÖHLER. Variance analysis for model updating with a finite element based subspace fitting approach, in "Mechanical Systems and Signal Processing", 2017, vol. 91, p. 142 - 156 [DOI: 10.1016/J.YMSSP.2017.01.006], https://hal.inria.fr/hal-01433175.
- [19] N. LE TOUZ, J. DUMOULIN, G. GENNARELLI, F. SOLDOVIERIA joint thermal and electromagnetic diagnostics approach for the inspection of thick walls, in "Geoscientific Instrumentation, Methods and Data

Systems Discussions", 2017, vol. 6, n<sup>o</sup> 1, p. 81 - 92 [*DOI* : 10.5194/GI-6-81-2017], https://hal.inria.fr/hal-01563859.

- [20] M. MARCHETTI, M. FOIS, L. IBOS, J. DUMOULIN, P. BOURSON, J. M. PIAU. Comparative study in the identification of liquid to solid transition phase with DSC, Raman spectra analysis and chemiometrics methods applied to phase change materials used for icing-delay in civil engineering infrastructures, in "Applied Thermal Engineering", 2017, https://hal-centralesupelec.archives-ouvertes.fr/hal-01629608.
- [21] Q. ZHANG, L. LJUNG.From Structurally Independent Local LTI Models to LPV Model, in "Automatica", August 2017, vol. 84, p. 232-235 [DOI : 10.1016/J.AUTOMATICA.2017.06.006], https://hal.inria.fr/hal-01635096.
- [22] Q. ZHANG. On Stability of the Kalman Filter for Discrete Time Output Error Systems, in "Systems and Control Letters", July 2017, vol. 107, p. 84-91 [DOI: 10.1016/J.SYSCONLE.2017.07.011], https://hal.inria.fr/hal-01635092.

#### **Invited Conferences**

[23] J. DUMOULIN.Infrared Thermography in Civil Engineering: From Non Destructive Testing in Laboratory to Outdoor Thermal Monitoring, in "QIRT-Asia 2017 - 2nd Asian Conference on Quantitative InfraRed Thermography", Daejeon, South Korea, July 2017, https://hal.inria.fr/hal-01651932.

#### **International Conferences with Proceedings**

- [24] S. ALLAHDADIAN, M. DÖHLER, C. VENTURA, L. MEVEL. Damage localization of a real structure using the statistical subspace damage localization method, in "IWSHM - 11th International Workshop on Structural Health Monitoring", Stanford, United States, September 2017, https://hal.inria.fr/hal-01589549.
- [25] N. BERRABAH, Q. ZHANG, M. FRANCHET, D. VAUTRIN. Estimation of a Cable Resistance Profile with Readaptation of Mismatched Measurement Instrument, in "International Conference on Electromagnetics in Advanced Applications (ICEAA 2017)", Verona, Italy, September 2017, https://hal.inria.fr/hal-01592434.
- [26] M. D. H. BHUYAN, S. ALLAHDADIAN, M. DÖHLER, Y. LECIEUX, L. MEVEL, F. SCHOEFS, C. VENTURA. Transfer matrices-based output-only statistical damage localization and quantification on a frame structure, in "IWSHM 11th International Workshop on Structural Health Monitoring", Stanford, United States, September 2017, https://hal.inria.fr/hal-01589536.
- [27] M. D. H. BHUYAN, G. GAUTIER, M. DÖHLER, Y. LECIEUX, L. MEVEL, F. SCHOEFS. Statistical damage localization in mechanical systems based on load vectors, in "IFAC WC - 20th IFAC World Congress", Toulouse, France, July 2017, https://hal.inria.fr/hal-01564490.
- [28] M. D. H. BHUYAN, E. VIEFHUES, M. DÖHLER, Y. LECIEUX, L. MEVEL, F. HILLE, F. SCHOEFS. *Output-only subspace and transfer matrix-based damage localization and quantification*, in "IOMAC 7th International Operational Modal Analysis Conference", Ingolstadt, Germany, May 2017, https://hal.inria.fr/hal-01522138.

#### [29] Best Paper

J. DUMOULIN, A. CRINIÈRE. *Infrared Thermography applied to transport infrastructures monitoring: outcomes and perspectives*, in "SPIE - Thermosense: Thermal Infrared Applications XXXIX", Anaheim California, United States, April 2017, https://hal.inria.fr/hal-01509555.

#### [30] Best Paper

M. DÖHLER, P. ANDERSEN, L. MEVEL. *Variance computation of modal parameter estimates from UPC subspace identification*, in "IOMAC - 7th International Operational Modal Analysis Conference", Ingolstadt, Germany, May 2017, https://hal.inria.fr/hal-01522137.

- [31] M. DÖHLER, Q. ZHANG, L. MEVEL. Change detection and isolation in mechanical system parameters based on perturbation analysis, in "IFAC WC - 20th IFAC World Congress", Toulouse, France, July 2017, https:// hal.inria.fr/hal-01564494.
- [32] S. GRES, M. DALGAARD ULRIKSEN, M. DÖHLER, R. J. JOHANSEN, P. ANDERSEN, L. DAMKILDE, S. A. NIELSEN. Statistical methods for damage detection applied to civil structures, in "EURODYN 2017 10th International Conference on Structural Dynamics", Rome, Italy, September 2017, vol. 199, p. 1919 1924 [DOI: 10.1016/J.PROENG.2017.09.280], https://hal.inria.fr/hal-01619442.
- [33] V. LE CAM, A. BOUCHE, D. PALLIER. Wireless Sensors Synchronization : an accurate and deterministic GPS-based algorithm, in "IWSHM 2017 - 11th international workshop on structural health monitoring 2017", Stanford, United States, September 2017, https://hal.inria.fr/hal-01633693.
- [34] N. LE TOUZ, J. DUMOULIN, J.-M. PIAU.Étude numérique de la résolution du couplage convection/radiation/diffusion dans une structure de chaussée hybride, in "25eme congrès français de thermique", Marseille, France, Société Française de Thermique (SFT), May 2017, https://hal.inria.fr/hal-01563862.
- [35] N. LE TOUZ, T. TOULLIER, J. DUMOULIN. Infrared thermography applied to the study of heated and solar pavement: from numerical modeling to small scale laboratory experiments, in "SPIE - Thermosense: Thermal Infrared Applications XXXIX", Anaheim, United States, April 2017, https://hal.inria.fr/hal-01563851.
- [36] L. LONG, S. THÖNS, M. DÖHLER. Damage Detection and Deteriorating Structural Systems, in "IWSHM - 11th International Workshop on Structural Health Monitoring", Stanford, United States, September 2017, https://hal.inria.fr/hal-01589545.
- [37] M. POURASGHAR, V. PUIG, C. OCAMPO-MARTINEZ, Q. ZHANG. Reduced-order Interval-observer Design for Dynamic Systems with Time-invariant Uncertainty, in "IFAC 2017 - 20th World Congress of the International Federation of Automatic Control", Toulouse, France, July 2017, p. 1-6, https://hal.inria.fr/hal-01635124.
- [38] S. SEN, A. CRINIÈRE, L. MEVEL, F. CÉROU, J. DUMOULIN. Seismic induced damage detection through parallel estimation of force and parameter using improved interacting Particle-Kalman filter, in "11th International Workshop on Structural Health Monitoring", San Francisco, United States, September 2017, https://hal.inria.fr/hal-01590713.

- [39] D. STRAUB, E. CHATZI, E. BISMUT, W. COURAGE, M. DÖHLER, M. H. FABER, J. KÖHLER, G. LOMBAERT, P. OMENZETTER, M. POZZI, S. THÖNS, D. V. VAL, H. WENZEL, D. ZONTA. Value of information: A roadmap to quantifying the benefit of structural health monitoring, in "ICOSSAR 12th International Conference on Structural Safety & Reliability", Vienna, Austria, August 2017, https://hal.inria. fr/hal-01577257.
- [40] S. THÖNS, M. P. LIMONGELLI, A. MANDIC IVANKOVIC, M. FABER, D. VAL, M. CHRYSSANTHOPLOUS, G. LOMBAERT, M. DÖHLER, D. STRAUB, E. CHATZI, J. KÖHLER, H. WENZEL, J. SØRENSEN. Progress of the COST Action TU1402 on the Quantification of the Value of Structural Health Monitoring, in "IWSHM - 11th International Workshop on Structural Health Monitoring", Stanford, United States, September 2017, https://hal.inria.fr/hal-01589527.
- [41] E. VIEFHUES, M. DÖHLER, F. HILLE, L. MEVEL. Stochastic Subspace-Based Damage Detection with Uncertainty in the Reference Null Space, in "IWSHM - 11th International Workshop on Structural Health Monitoring", Stanford, United States, September 2017, https://hal.inria.fr/hal-01589540.
- [42] Y. YANG, A. SEMPEY, T. VOGT-WU, A. SOMMIER, J. DUMOULIN, C. PRADERE, J.-C. BAT-SALE.Study of methodology for quantitative thermal diagnostic of wall, in "QIRT-Asia 2017 - 2nd Asian Conference on Quantitative InfraRed Thermography", Daejeon, South Korea, July 2017, p. 1-3 [DOI: 10.21611/QIRT.2017.014], https://hal.inria.fr/hal-01651857.
- [43] Y. YANG, T. VOGT WU, A. SEMPEY, J. DUMOULIN, J. C. BATSALE.Study Of Experimental Diagnostic Methodology For Thermal Characterization Of Building Wall, in "International Conference on Materials & energy", Tianjin, China, July 2017, https://hal.inria.fr/hal-01670127.
- [44] Q. ZHANG. Adaptive Kalman Filter for Actuator Fault Diagnosis, in "IFAC 2017 20th World Congress of the International Federation of Automatic Control", Toulouse, France, July 2017, p. 1-6, https://hal.inria.fr/hal-01635108.
- [45] Q. ZHANG. Inverse Scattering for Electrical Cable Soft Fault Diagnosis, in "IFAC 2017 20th World Congress of the International Federation of Automatic Control", Toulouse, France, July 2017, p. 1-2, https://hal.inria.fr/ hal-01635131.

#### **National Conferences with Proceeding**

- [46] A. CRINIÈRE, J. DUMOULIN, L. MEVEL. Surveillance thermique de structures par thermographie infrarouge couplée au suivi des paramètres environnementaux : du déploiement à la standardisation, in "Folle journée de l'imagerie Nantaise 2017", Nantes, France, February 2017, https://hal.inria.fr/hal-01493289.
- [47] A. CRINIÈRE, A. ROUMY, T. MAUGEY, M. KIEFFER, J. DUMOULIN. Sélection optimale de capteurs de référence pour le stockage de données spatialement corrélées, in "GRETSI 2017 - XXVIème Colloque", Juan-les-Pins, France, GRETSI 2017, September 2017, https://hal.inria.fr/hal-01575610.
- [48] T. TOULLIER, J. DUMOULIN, L. MEVEL. Étude comparative de deux approches, thermocouples intégrés et thermographie infrarouge, pour la surveillance thermique d'une infrastructure de transport, in "25eme congrès français de thermique", Marseille, France, SFT, May 2017, https://hal.inria.fr/hal-01563842.

#### **Conferences without Proceedings**

[49] S. SEN, A. CRINIÈRE, L. MEVEL, F. CÉROU, J. DUMOULIN. Estimation of time varying system parameters from ambient response using improved Particle-Kalman filter with correlated noise, in "EGU General Assembly 2017", Vienne, Austria, April 2017, https://hal.inria.fr/hal-01493840.

#### Scientific Books (or Scientific Book chapters)

- [50] J. DUMOULIN.Infrared thermography: from sensing principle to non destructive testing considerations, in "Sensing the past", Geoscience and Sensing Technologies for cultural heritage, Springer, 2017, https://hal. inria.fr/hal-01376217.
- [51] F. SOLDOVIERI, J. DUMOULIN. Integrated monitoring at a modern architectural masterpiece: the case of Viaduct Basento in Potenza, in "Sensing the past", Geoscience and Sensing Technologies for cultural heritage, Springer, 2017, n<sup>o</sup> 16, p. 499-514, https://hal.inria.fr/hal-01376212.

#### **Other Publications**

- [52] G. GAUTIER, M. D. H. BHUYAN, M. DÖHLER, L. MEVEL. Comparison of damage localization in mechanical systems based on Stochastic Subspace Identification method, April 2017, EGU General Assembly, Poster, https://hal.inria.fr/hal-01545346.
- [53] M. MARCHETTI, J. DUMOULIN, L. IBOS, J. M. PIAU. Coupling Raman spectroscopy and infrared thermography to evaluate energy exchanges with PCM embedded into pavement, April 2017, European Geosciences Union General Assembly 2017, Poster, https://hal-centralesupelec.archives-ouvertes.fr/hal-01629617.
- [54] Y. YANG, A. SEMPEY, T. VOGT WU, A. SOMMIER, J. DUMOULIN, J. C. BATSALE. Study of an experimental methodology for thermal properties diagnostic of building envelop, April 2017, 1, EGU 2017 - European Geosciences Union General Assembly, Poster, https://hal.inria.fr/hal-01651660.

### **References in notes**

- [55] M. BASSEVILLE, I. V. NIKIFOROV. Fault isolation for diagnosis : nuisance rejection and multiple hypotheses testing, in "Annual Reviews in Control", December 2002, vol. 26, n<sup>o</sup> 2, p. 189–202, http://dx.doi.org/10.1016/ S1367-5788(02)00029-9.
- [56] B. DELYON, A. JUDITSKY, A. BENVENISTE. On the relationship between identification and local tests, IRISA, May 1997, n<sup>o</sup> 1104, ftp://ftp.irisa.fr/techreports/1997/PI-1104.ps.gz.
- [57] M. JAULENT. *The inverse scattering problem for LCRG transmission lines*, in "Journal of Mathematical Physics", December 1982, vol. 23, n<sup>o</sup> 12, p. 2286-2290.
- [58] G. L. LAMB. Elements of Soliton Theory, John Wiley & Sons, New York, 1980.
- [59] M. OUMRI.*Fault diagnosis of wired electric networks by reflectometry*, Université Paris Sud Paris XI, May 2014, https://tel.archives-ouvertes.fr/tel-01165039.
- [60] C. R. PAUL. Analysis of multiconductor transmission lines, Wiley, New York, 2008.

- [61] H. TANG, Q. ZHANG. An Inverse Scattering Approach to Soft Fault Diagnosis in Lossy Electric Transmission Lines, in "IEEE Trans. on Antennas and Propagation", 2011, vol. 59, n<sup>o</sup> 10, p. 3730-3737, http://dx.doi.org/ 10.1109/TAP.2011.2163772.
- [62] P. VAN OVERSCHEE, B. DE MOOR. *Subspace Identification for Linear Systems*, Kluwer Academic Publishers, Boston, 1996.
- [63] F. VISCO COMANDINI.Some inverse scattering problems on star-shaped graphs: application to fault detection on electrical transmission line networks, Université de Versailles-Saint Quentin en Yvelines, December 2011, https://tel.archives-ouvertes.fr/tel-00748216.

# **Project-Team IPSO**

# **Invariant Preserving SOlvers**

IN COLLABORATION WITH: Institut de recherche mathématique de Rennes (IRMAR)

IN PARTNERSHIP WITH: CNRS Université Rennes 1 École normale supérieure de Rennes

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Numerical schemes and simulations

# **Table of contents**

1.	Personnel	481
2.	Overall Objectives	482
	2.1. An overview of geometric numerical integration	482
	2.2. Overall objectives	482
3.	Research Program	483
	3.1. Structure-preserving numerical schemes for solving ordinary differential equations	483
	3.1.1. Reversible ODEs	483
	3.1.2. ODEs with an invariant manifold	483
	3.1.3. Hamiltonian systems	484
	3.1.4. Differential-algebraic equations	485
	3.2. Highly-oscillatory systems	485
	3.3. Geometric schemes for the Schrödinger equation	486
	3.4. High-frequency limit of the Helmholtz equation	487
	3.5. From the Schrödinger equation to Boltzmann-like equations	487
4.	New Results	488
	4.1. Multiscale numerical methods	488
	4.1.1. Asymptotic preserving and time diminishing schemes for rarefied gas dynamic	488
	4.1.2. An exponential integrator for the drift-kinetic model	489
	4.1.3. Multiscale Particle-in-Cell methods and comparisons for the long-time two-dimensional states and the states of the long-time two-dimensional states and the states of the long-time two-dimensional states are stated as the states of the long-time two-dimensional states are stated as the states of the long-time two-dimensional states are stated as the states of the long-time two-dimensional states are stated as the long-time two-dimensional states are states are states are stated as the long-time two-dimensional states are states a	sional
	Vlasov-Poisson equation with strong magnetic field	489
	4.1.4. Nonlinear Geometric Optics based multiscale stochastic Galerkin methods for h	nighly
	oscillatory transport equations with random inputs	489
	4.1.5. Nonlinear Geometric Optics method based multi-scale numerical schemes for hi	Igniy-
	4.1.6 Uish and a Hamiltonian anlitting for Vlacon Deison constinue	489
	4.1.0. High-order Hammonian spinting for viasov-Poisson equations	490
	4.1.7. A particle inicio-macio decomposition based numerical scheme for confisional k	111euc 400
	4.1.8 Uniformly accurate forward semi Lagrangian methods for highly oscillatory VI	490
	Poisson equation	
	4.1.9 Uniformly accurate multiscale time integrators for second order oscillatory differ	ential
	equations with large initial data	490
	4.1.10. Unconditional and optimal $H^2$ -error estimates of two linear and conservative	finite
	difference schemes for the Klein-Gordon-Schrödinger equation in high dimensions	491
	4.1.11. A combination of multiscale time integrator and two-scale formulation for the non	linear
	Schrödinger equation with wave operator	491
	4.1.12. Uniformly accurate numerical schemes for the nonlinear Dirac equation in the non	nrela-
	tivistic limit regime	491
	4.1.13. A formal series approach to the center manifold theorem	491
	4.1.14. Convergence of multi-revolution composition time-splitting methods for highly oscil	latory
	differential equations of Schrödinger type	492
	4.1.15. Highly-oscillatory evolution equations with multiple frequencies: averaging and num	nerics
		492
	4.1.16. Optimality and resonances in a class of compact finite difference schemes of high or	rder492
	4.2. mathematical analysis of multiscale partial differential equations	492
	4.2.1. Collision of almost parallel vortex filaments	492
	4.2.2. Free vibrations of axisymmetric shells: parabolic and elliptic cases	492
	4.2.3. High frequency oscillations of first eigenmodes in axisymmetric shells as the thic	kness
	tends to zero	493
	4.2.4. Semiclassical Sobolev constants for the electro-magnetic Robin Laplacian	493

	4.2.5.	On the MIT Bag Model in the Non-relativistic Limit	493
	4.2.6.	Dimension reduction for dipolar Bose-Einstein condensates in the strong inter-	action
	1	regime	493
	4.2.7.	Nonlinear stability criteria for the HMF Model	493
	4.2.8.	Strong confinement limit for the nonlinear Schrödinger equation constrained on a cu	ırve <mark>493</mark>
	4.2.9.	Stable ground states for the HMF Poisson Model	494
	4.2.10.	The quantum Liouville-BGK equation and the moment problem	494
	4.2.11.	Averaging of nonlinear Schrödinger equations with strong magnetic confinement	494
	4.3. ma	thematical analysis of stochastic partial differential equations	494
	4.3.1.	Large deviations for the dynamic $\Phi_d^{2n}$ model	494
	4.3.2.	Solution to the stochastic Schrödinger equation on the full space	494
	4.3.3.	A law of large numbers in the supremum norm for a multiscale stochastic spatial	gene
	1	network	495
	4.3.4.	Long time behavior of Gross-Pitaevskii equation at positive temperature	495
	4.3.5.	An integral inequality for the invariant measure of a stochastic reaction-diffusion equ	uation
			495
	4.3.6.	Kolmogorov equations and weak order analysis for SPDES with nonlinear diff	fusion
		coefficient	495
	4.3.7.	Approximation-diffusion in stochastically forced kinetic equations	495
5.	Partnersh	ips and Cooperations	495
	5.1. Reg	gional Initiatives	495
	5.2. Na	tional Initiatives	496
	5.2.1.	ANR MOONRISE: 2015-2019	496
	5.2.2.	ANR MFG: 2016-2020	496
	5.2.3.	ANR ACHYLLES: 2014-2018	496
	5.2.4.	IPL FRATRES	497
	5.3. Eu	ropean Initiatives	497
	5.4. Inte	ernational Initiatives	497
	5.4.1.	Informal International Partners	497
	5.4.2.	Participation in Other International Programs	498
(	5.5. Into	ernational Research Visitors	498
0.	Dissemina		498
	0.1. Pro	Scientific Exercise	498
	0.1.1.	Scientific Events Organisation	498
	0.1.2.	JOUINAI	498
	0.1	2.2. Deviewer Deviewing Activities	498
	612	Levited Telles	498
	0.1.3.	nivited faiks	499
	0.1.4. 6 1 5	Scientific Expertise Desearch Administration	500
	6.2 Tar	Research Administration	500
	6.2.1 102	Teaching	500
	622	Supervision	500
	622		501
7	0.2.3. Bibliogra	nhv	502
<i>'</i> •	Dibilogra	Ридн.	

# **Project-Team IPSO**

*Creation of the Project-Team: 2004 December 06, end of the Project-Team: 2017 December 31* **Keywords:** 

### **Computer Science and Digital Science:**

A6. - Modeling, simulation and control

A6.1. - Mathematical Modeling

A6.1.1. - Continuous Modeling (PDE, ODE)

- A6.1.2. Stochastic Modeling (SPDE, SDE)
- A6.1.4. Multiscale modeling

A6.1.5. - Multiphysics modeling

A6.2. - Scientific Computing, Numerical Analysis & Optimization

A6.2.1. - Numerical analysis of PDE and ODE

A6.2.2. - Numerical probability

A6.2.3. - Probabilistic methods

#### **Other Research Topics and Application Domains:**

B1. - Life sciences

B1.1. - Biology

B1.1.10. - Mathematical biology

B4. - Energy

B4.1. - Fossile energy production (oil, gas)

B5. - Industry of the future

B5.3. - Nanotechnology

# 1. Personnel

#### **Research Scientists**

Philippe Chartier [Team leader, Inria, Senior Researcher, HDR] Nicolas Crouseilles [Inria, Researcher, HDR] Erwan Faou [Inria, Senior Researcher, HDR] Mohammed Lemou [CNRS, Senior Researcher, HDR]

#### **Faculty Members**

Francois Castella [Univ de Rennes I, Professor, HDR] Anais Crestetto [Univ de Nantes, Associate Professor] Arnaud Debussche [Ecole normale supérieure de Rennes, Professor, HDR] Florian Mehats [Univ de Rennes I, Professor, HDR]

#### **PhD Students**

Joackim Bernier [Univ de Rennes I] Marine Fontaine [Ecole normale supérieure de Rennes] Romain Horsin [Inria, until Nov 2017] Angelo Rosello [Ecole normale supérieure de Rennes, from Sep 2017]

#### **Post-Doctoral Fellows**

Nathalie Ayi [Inria, until Aug 2017] Xiaofei Zhao [Inria, until Sep 2017] Administrative Assistant Stephanie Lemaile [Inria]

# 2. Overall Objectives

## 2.1. An overview of geometric numerical integration

A fundamental and enduring challenge in science and technology is the quantitative prediction of timedependent nonlinear phenomena. While dynamical simulation (for ballistic trajectories) was one of the first applications of the digital computer, the problems treated, the methods used, and their implementation have all changed a great deal over the years. Astronomers use simulation to study long term evolution of the solar system. Molecular simulations are essential for the design of new materials and for drug discovery. Simulation can replace or guide experiment, which often is difficult or even impossible to carry out as our ability to fabricate the necessary devices is limited.

During the last decades, we have seen dramatic increases in computing power, bringing to the fore an ever widening spectrum of applications for dynamical simulation. At the boundaries of different modeling regimes, it is found that computations based on the fundamental laws of physics are under-resolved in the textbook sense of numerical methods. Because of the vast range of scales involved in modeling even relatively simple biological or material functions, this limitation will not be overcome by simply requiring more computing power within any realistic time. One therefore has to develop numerical methods which capture crucial structures even if the method is far from "converging" in the mathematical sense. In this context, we are forced increasingly to think of the numerical algorithm as a part of the modeling process itself. A major step forward in this area has been the development of structure-preserving or "geometric" integrators which maintain conservation laws, dissipation rates, or other key features of the continuous dynamical model. Conservation of energy and momentum are fundamental for many physical models; more complicated invariants are maintained in applications such as molecular dynamics, vehicle simulation, astrodynamics) the available structure may include constraint dynamics, actuator or thruster geometry, dissipation rates and properties determined by nonlinear forms of damping.

In recent years the growth of geometric integration has been very noticeable. Features such as *symplecticity* or *time-reversibility* are now widely recognized as essential properties to preserve, owing to their physical significance. This has motivated a lot of research [42], [39], [38] and led to many significant theoretical achievements (symplectic and symmetric methods, volume-preserving integrators, Lie-group methods, ...). In practice, a few simple schemes such as the Verlet method or the Störmer method have been used for years with great success in molecular dynamics or astronomy. However, they now need to be further improved in order to fit the tremendous increase of complexity and size of the models.

### 2.2. Overall objectives

To become more specific, the project *IPSO* aims at finding and implementing new structure-preserving schemes and at understanding the behavior of existing ones for the following type of problems:

- systems of differential equations posed on a manifold.
- systems of differential-algebraic equations of index 2 or 3, where the constraints are part of the equations.
- Hamiltonian systems and constrained Hamiltonian systems (which are special cases of the first two items though with some additional structure).
- highly-oscillatory systems (with a special focus of those resulting from the Schrödinger equation).

Although the field of application of the ideas contained in geometric integration is extremely wide (e.g. robotics, astronomy, simulation of vehicle dynamics, biomechanical modeling, biomolecular dynamics, geodynamics, chemistry...), *IPSO* will mainly concentrate on applications for *molecular dynamics simulation* and *laser simulation*:

- There is a large demand in biomolecular modeling for models that integrate microscopic molecular dynamics simulation into statistical macroscopic quantities. These simulations involve huge systems of ordinary differential equations over very long time intervals. This is a typical situation where the determination of accurate trajectories is out of reach and where one has to rely on the good qualitative behavior of structure-preserving integrators. Due to the complexity of the problem, more efficient numerical schemes need to be developed.
- The demand for new models and/or new structure-preserving schemes is also quite large in laser simulations. The propagation of lasers induces, in most practical cases, several well-separated scales: the intrinsically highly-oscillatory *waves* travel over long distances. In this situation, filtering the oscillations in order to capture the long-term trend is what is required by physicists and engineers.

# 3. Research Program

# 3.1. Structure-preserving numerical schemes for solving ordinary differential equations

Participants: Francois Castella, Philippe Chartier, Erwan Faou.

ordinary differential equation, numerical integrator, invariant, Hamiltonian system, reversible system, Liegroup system

In many physical situations, the time-evolution of certain quantities may be written as a Cauchy problem for a differential equation of the form

$$y'(t) = f(y(t)),$$
  
 $y(0) = y_0.$ 
(38)

For a given  $y_0$ , the solution y(t) at time t is denoted  $\varphi_t(y_0)$ . For fixed t,  $\varphi_t$  becomes a function of  $y_0$  called the flow of (1). From this point of view, a numerical scheme with step size h for solving (1) may be regarded as an approximation  $\Phi_h$  of  $\varphi_h$ . One of the main questions of geometric integration is whether intrinsic properties of  $\varphi_t$  may be passed on to  $\Phi_h$ .

This question can be more specifically addressed in the following situations:

#### 3.1.1. Reversible ODEs

The system (1) is said to be  $\rho$ -reversible if there exists an involutive linear map  $\rho$  such that

$$\rho \circ \varphi_t = \varphi_t^{-1} \circ \rho = \varphi_{-t} \circ \rho. \tag{39}$$

It is then natural to require that  $\Phi_h$  satisfies the same relation. If this is so,  $\Phi_h$  is said to be *symmetric*. Symmetric methods for reversible systems of ODEs are just as much important as *symplectic* methods for Hamiltonian systems and offer an interesting alternative to symplectic methods.

#### 3.1.2. ODEs with an invariant manifold

The system (1) is said to have an invariant manifold g whenever

$$\mathcal{M} = \{ y \in \mathbb{R}^n; g(y) = 0 \}$$
(40)

is kept *globally* invariant by  $\varphi_t$ . In terms of derivatives and for sufficiently differentiable functions f and g, this means that

$$\forall y \in \mathcal{M}, g'(y)f(y) = 0.$$

As an example, we mention Lie-group equations, for which the manifold has an additional group structure. This could possibly be exploited for the space-discretisation. Numerical methods amenable to this sort of problems have been reviewed in a recent paper [37] and divided into two classes, according to whether they use g explicitly or through a projection step. In both cases, the numerical solution is forced to live on the manifold at the expense of some Newton's iterations.

#### 3.1.3. Hamiltonian systems

Hamiltonian problems are ordinary differential equations of the form:

$$\dot{p}(t) = -\nabla_q H(p(t), q(t)) \in \mathbb{R}^d$$

$$\dot{q}(t) = \nabla_p H(p(t), q(t)) \in \mathbb{R}^d$$

$$(41)$$

with some prescribed initial values  $(p(0), q(0)) = (p_0, q_0)$  and for some scalar function H, called the Hamiltonian. In this situation, H is an invariant of the problem. The evolution equation (4) can thus be regarded as a differential equation on the manifold

$$\mathcal{M} = \{ (p,q) \in \mathbb{R}^d \times \mathbb{R}^d; H(p,q) = H(p_0,q_0) \}.$$

Besides the Hamiltonian function, there might exist other invariants for such systems: when there exist d invariants in involution, the system (4) is said to be *integrable*. Consider now the parallelogram P originating from the point  $(p, q) \in \mathbb{R}^{2d}$  and spanned by the two vectors  $\xi \in \mathbb{R}^{2d}$  and  $\eta \in \mathbb{R}^{2d}$ , and let  $\omega(\xi, \eta)$  be the sum of the *oriented* areas of the projections over the planes  $(p_i, q_i)$  of P,

$$\omega(\xi,\eta) = \xi^T J\eta_z$$

where J is the *canonical symplectic* matrix

$$J = \left[ \begin{array}{cc} 0 & I_d \\ -I_d & 0 \end{array} \right].$$

A continuously differentiable map g from  $\mathbb{R}^{2d}$  to itself is called symplectic if it preserves  $\omega$ , i.e. if

$$\omega(g'(p,q)\xi,g'(p,q)\eta) = \omega(\xi,\eta).$$

A fundamental property of Hamiltonian systems is that their exact flow is symplectic. Integrable Hamiltonian systems behave in a very remarkable way: as a matter of fact, their invariants persist under small perturbations, as shown in the celebrated theory of Kolmogorov, Arnold and Moser. This behavior motivates the introduction of *symplectic* numerical flows that share most of the properties of the exact flow. For practical simulations of Hamiltonian systems, symplectic methods possess an important advantage: the error-growth as a function of time is indeed linear, whereas it would typically be quadratic for non-symplectic methods.

#### 3.1.4. Differential-algebraic equations

Whenever the number of differential equations is insufficient to determine the solution of the system, it may become necessary to solve the differential part and the constraint part altogether. Systems of this sort are called differential-algebraic systems. They can be classified according to their index, yet for the purpose of this expository section, it is enough to present the so-called index-2 systems

$$\dot{y}(t) = f(y(t), z(t)), 
0 = g(y(t)),$$
(42)

where initial values  $(y(0), z(0)) = (y_0, z_0)$  are given and assumed to be consistent with the constraint manifold. By constraint manifold, we imply the intersection of the manifold

$$\mathcal{M}_1 = \{ y \in \mathbb{R}^n, g(y) = 0 \}$$

and of the so-called hidden manifold

$$\mathcal{M}_2 = \{(y, z) \in \mathbb{R}^n \times \mathbb{R}^m, \frac{\partial g}{\partial y}(y)f(y, z) = 0\}.$$

This manifold  $\mathcal{M} = \mathcal{M}_1 \cap \mathcal{M}_2$  is the manifold on which the exact solution (y(t), z(t)) of (5) lives.

There exists a whole set of schemes which provide a numerical approximation lying on  $\mathcal{M}_1$ . Furthermore, this solution can be projected on the manifold  $\mathcal{M}$  by standard projection techniques. However, it it worth mentioning that a projection destroys the symmetry of the underlying scheme, so that the construction of a symmetric numerical scheme preserving  $\mathcal{M}$  requires a more sophisticated approach.

#### **3.2. Highly-oscillatory systems**

**Participants:** Francois Castella, Philippe Chartier, Nicolas Crouseilles, Erwan Faou, Florian Mehats, Mohammed Lemou.

second-order ODEs, oscillatory solutions, Schrödinger and wave equations, step size restrictions.

In applications to molecular dynamics or quantum dynamics for instance, the right-hand side of (1) involves *fast* forces (short-range interactions) and *slow* forces (long-range interactions). Since *fast* forces are much cheaper to evaluate than *slow* forces, it seems highly desirable to design numerical methods for which the number of evaluations of slow forces is not (at least not too much) affected by the presence of fast forces.

A typical model of highly-oscillatory systems is the second-order differential equations

$$\ddot{q} = -\nabla V(q) \tag{43}$$

where the potential V(q) is a sum of potentials V = W + U acting on different time-scales, with  $\nabla^2 W$  positive definite and  $\|\nabla^2 W\| >> \|\nabla^2 U\|$ . In order to get a bounded error propagation in the linearized equations for an explicit numerical method, the step size must be restricted according to

$$h\omega < C,$$

where C is a constant depending on the numerical method and where  $\omega$  is the highest frequency of the problem, i.e. in this situation the square root of the largest eigenvalue of  $\nabla^2 W$ . In applications to molecular dynamics for instance, *fast* forces deriving from W (short-range interactions) are much cheaper to evaluate than *slow* forces deriving from U (long-range interactions). In this case, it thus seems highly desirable to design numerical methods for which the number of evaluations of slow forces is not (at least not too much) affected by the presence of fast forces. Another prominent example of highly-oscillatory systems is encountered in quantum dynamics where the Schrödinger equation is the model to be used. Assuming that the Laplacian has been discretized in space, one indeed gets the *time*-dependent Schrödinger equation:

$$i\dot{\psi}(t) = \frac{1}{\varepsilon}H(t)\psi(t),\tag{44}$$

where H(t) is finite-dimensional matrix and where  $\varepsilon$  typically is the square-root of a mass-ratio (say electron/ion for instance) and is small ( $\varepsilon \approx 10^{-2}$  or smaller). Through the coupling with classical mechanics (H(t) is obtained by solving some equations from classical mechanics), we are faced once again with two different time-scales, 1 and  $\varepsilon$ . In this situation also, it is thus desirable to devise a numerical method able to advance the solution by a time-step  $h > \varepsilon$ .

# 3.3. Geometric schemes for the Schrödinger equation

Participants: Francois Castella, Philippe Chartier, Erwan Faou, Florian Mehats.

Schrödinger equation, variational splitting, energy conservation.

Given the Hamiltonian structure of the Schrödinger equation, we are led to consider the question of energy preservation for time-discretization schemes.

At a higher level, the Schrödinger equation is a partial differential equation which may exhibit Hamiltonian structures. This is the case of the time-dependent Schrödinger equation, which we may write as

$$i\varepsilon\frac{\partial\psi}{\partial t} = H\psi,\tag{45}$$

where  $\psi = \psi(x, t)$  is the wave function depending on the spatial variables  $x = (x_1, \dots, x_N)$  with  $x_k \in \mathbb{R}^d$ (e.g., with d = 1 or 3 in the partition) and the time  $t \in \mathbb{R}$ . Here,  $\varepsilon$  is a (small) positive number representing the scaled Planck constant and i is the complex imaginary unit. The Hamiltonian operator H is written

$$H = T + V$$

with the kinetic and potential energy operators

$$T = -\sum_{k=1}^{N} \frac{\varepsilon^2}{2m_k} \Delta_{x_k}$$
 and  $V = V(x)$ ,

where  $m_k > 0$  is a particle mass and  $\Delta_{x_k}$  the Laplacian in the variable  $x_k \in \mathbb{R}^d$ , and where the real-valued potential V acts as a multiplication operator on  $\psi$ .

The multiplication by i in (8) plays the role of the multiplication by J in classical mechanics, and the energy  $\langle \psi | H | \psi \rangle$  is conserved along the solution of (8), using the physicists' notations  $\langle u | A | u \rangle = \langle u, Au \rangle$  where  $\langle , \rangle$  denotes the Hermitian  $L^2$ -product over the phase space. In quantum mechanics, the number N of particles is very large making the direct approximation of (8) very difficult.

The numerical approximation of (8) can be obtained using projections onto submanifolds of the phase space, leading to various PDEs or ODEs: see [41], [40] for reviews. However the long-time behavior of these approximated solutions is well understood only in this latter case, where the dynamics turns out to be finite dimensional. In the general case, it is very difficult to prove the preservation of qualitative properties of (8) such as energy conservation or growth in time of Sobolev norms. The reason for this is that backward error analysis is not directly applicable for PDEs. Overwhelming these difficulties is thus a very interesting challenge.

A particularly interesting case of study is given by symmetric splitting methods, such as the Strang splitting:

$$\psi_1 = \exp\left(-i(\delta t)V/2\right)\exp\left(i(\delta t)\Delta\right)\exp\left(-i(\delta t)V/2\right)\psi_0\tag{46}$$

where  $\delta t$  is the time increment (we have set all the parameters to 1 in the equation). As the Laplace operator is unbounded, we cannot apply the standard methods used in ODEs to derive long-time properties of these schemes. However, its projection onto finite dimensional submanifolds (such as Gaussian wave packets space or FEM finite dimensional space of functions in x) may exhibit Hamiltonian or Poisson structure, whose long-time properties turn out to be more tractable.

### 3.4. High-frequency limit of the Helmholtz equation

Participant: Francois Castella.

waves, Helmholtz equation, high oscillations.

The Helmholtz equation models the propagation of waves in a medium with variable refraction index. It is a simplified version of the Maxwell system for electro-magnetic waves.

The high-frequency regime is characterized by the fact that the typical wavelength of the signals under consideration is much smaller than the typical distance of observation of those signals. Hence, in the high-frequency regime, the Helmholtz equation at once involves highly oscillatory phenomena that are to be described in some asymptotic way. Quantitatively, the Helmholtz equation reads

$$i\alpha_{\varepsilon}u_{\varepsilon}(x) + \varepsilon^{2}\Delta_{x}u_{\varepsilon} + n^{2}(x)u_{\varepsilon} = f_{\varepsilon}(x).$$
(47)

Here,  $\varepsilon$  is the small adimensional parameter that measures the typical wavelength of the signal, n(x) is the space-dependent refraction index, and  $f_{\varepsilon}(x)$  is a given (possibly dependent on  $\varepsilon$ ) source term. The unknown is  $u_{\varepsilon}(x)$ . One may think of an antenna emitting waves in the whole space (this is the  $f_{\varepsilon}(x)$ ), thus creating at any point x the signal  $u_{\varepsilon}(x)$  along the propagation. The small  $\alpha_{\varepsilon} > 0$  term takes into account damping of the waves as they propagate.

One important scientific objective typically is to describe the high-frequency regime in terms of *rays* propagating in the medium, that are possibly refracted at interfaces, or bounce on boundaries, etc. Ultimately, one would like to replace the true numerical resolution of the Helmholtz equation by that of a simpler, asymptotic model, formulated in terms of rays.

In some sense, and in comparison with, say, the wave equation, the specificity of the Helmholtz equation is the following. While the wave equation typically describes the evolution of waves between some initial time and some given observation time, the Helmholtz equation takes into account at once the propagation of waves over *infinitely long* time intervals. Qualitatively, in order to have a good understanding of the signal observed in some bounded region of space, one readily needs to be able to describe the propagative phenomena in the whole space, up to infinity. In other words, the "rays" we refer to above need to be understood from the initial time up to infinity. This is a central difficulty in the analysis of the high-frequency behaviour of the Helmholtz equation.

#### 3.5. From the Schrödinger equation to Boltzmann-like equations

Participant: Francois Castella.

Schrödinger equation, asymptotic model, Boltzmann equation.

The Schrödinger equation is the appropriate way to describe transport phenomena at the scale of electrons. However, for real devices, it is important to derive models valid at a larger scale. In semi-conductors, the Schrödinger equation is the ultimate model that allows to obtain quantitative information about electronic transport in crystals. It reads, in convenient adimensional units,

$$i\partial_t \psi(t,x) = -\frac{1}{2}\Delta_x \psi + V(x)\psi, \qquad (48)$$

where V(x) is the potential and  $\psi(t, x)$  is the time- and space-dependent wave function. However, the size of real devices makes it important to derive simplified models that are valid at a larger scale. Typically, one wishes to have kinetic transport equations. As is well-known, this requirement needs one to be able to describe "collisions" between electrons in these devices, a concept that makes sense at the macroscopic level, while it does not at the microscopic (electronic) level. Quantitatively, the question is the following: can one obtain the Boltzmann equation (an equation that describes collisional phenomena) as an asymptotic model for the Schrödinger equation, along the physically relevant micro-macro asymptotics? From the point of view of modelling, one wishes here to understand what are the "good objects", or, in more technical words, what are the relevant "cross-sections", that describe the elementary collisional phenomena. Quantitatively, the Boltzmann equation reads, in a simplified, linearized, form :

$$\partial_t f(t, x, v) = \int_{\mathbf{R}^3} \sigma(v, v') \left[ f(t, x, v') - f(t, x, v) \right] dv'.$$

$$\tag{49}$$

Here, the unknown is f(x, v, t), the probability that a particle sits at position x, with a velocity v, at time t. Also,  $\sigma(v, v')$  is called the cross-section, and it describes the probability that a particle "jumps" from velocity v to velocity v' (or the converse) after a collision process.

# 4. New Results

### 4.1. Multiscale numerical methods

#### 4.1.1. Asymptotic preserving and time diminishing schemes for rarefied gas dynamic

In [10], we introduce a new class of numerical schemes for rarefied gas dynamic problems described by collisional kinetic equations. The idea consists in reformulating the problem using a micro-macro decomposition and successively in solving the microscopic part by using asymptotic preserving Monte Carlo methods. We consider two types of decompositions, the first leading to the Euler system of gas dynamics while the second to the Navier-Stokes equations for the macroscopic part. In addition, the particle method which solves the microscopic part is designed in such a way that the global scheme becomes computationally less expensive as the solution approaches the equilibrium state as opposite to standard methods for kinetic equations which computational cost increases with the number of interactions. At the same time, the statistical error due to the particle part of the solution decreases as the system approach the equilibrium state. This causes the method to degenerate to the sole solution of the macroscopic hydrodynamic equations (Euler or Navier-Stokes) in the limit of infinite number of collisions. In a last part, we will show the behaviors of this new approach in comparisons to standard Monte Carlo techniques for solving the kinetic equation by testing it on different problems which typically arise in rarefied gas dynamic simulations.

#### 4.1.2. An exponential integrator for the drift-kinetic model

In [30], we propose an exponential integrator for the drift-kinetic equations in polar geometry. This approach removes the CFL condition from the linear part of the system (which is often the most stringent requirement in practice) and treats the remainder explicitly using Arakawa's finite difference scheme. The present approach is mass conservative, up to machine precision, and significantly reduces the computational effort per time step. In addition, we demonstrate the efficiency of our method by performing numerical simulations in the context of the ion temperature gradient instability. In particular, we find that our numerical method can take time steps comparable to what has been reported in the literature for the (predominantly used) splitting approach. In addition, the proposed numerical method has significant advantages with respect to conservation of energy and efficient higher order methods can be obtained easily. We demonstrate this by investigating the performance of a fourth order implementation.

### 4.1.3. Multiscale Particle-in-Cell methods and comparisons for the long-time two-dimensional Vlasov-Poisson equation with strong magnetic field

In [11], we applied different kinds of multiscale methods to numerically study the long-time Vlasov-Poisson equation with a strong magnetic field. The multiscale methods include an asymptotic preserving Runge-Kutta scheme, an exponential time differencing scheme, stroboscopic averaging method and a uniformly accurate two-scale formulation. We briefly review these methods and then adapt them to solve the Vlasov-Poisson equation under a Particle-in-Cell discretization. Extensive numerical experiments are conducted to investigate and compare the accuracy, efficiency, and long-time behavior of all the methods. The methods with the best performance under different parameter regimes are identified.

# 4.1.4. Nonlinear Geometric Optics based multiscale stochastic Galerkin methods for highly oscillatory transport equations with random inputs

In [31], we develop generalized polynomial chaos (gPC) based stochastic Galerkin (SG) methods for a class of highly oscillatory transport equations that arise in semiclassical modeling of non-adiabatic quantum dynamics. These models contain uncertainties, particularly in coefficients that correspond to the potentials of the molecular system. We first focus on a highly oscillatory scalar model with random uncertainty. Our method is built upon the nonlinear geometrical optics (NGO) based method, developed in [12] for numerical approximations of deterministic equations, which can obtain accurate pointwise solution even without numerically resolving spatially and temporally the oscillations. With the random uncertainty, we show that such a method has oscillatory higher order derivatives in the random space, thus requires a frequency dependent discretization in the random space. We modify this method by introducing a new " time " variable based on the phase, which is shown to be non-oscillatory in the random space, based on which we develop a gPC-SG method that can capture oscillations with the frequency-independent time step, mesh size as well as the degree of polynomial chaos. A similar approach is then extended to a semiclassical surface hopping model system with a similar numerical conclusion. Various numerical examples attest that these methods indeed capture accurately the solution statistics pointwisely even though none of the numerical parameters resolve the high frequencies of the solution.

## 4.1.5. Nonlinear Geometric Optics method based multi-scale numerical schemes for highly-oscillatory transport equations

In [12], we introduce a new numerical strategy to solve a class of oscillatory transport PDE models which is able to capture accurately the solutions without numerically resolving the high frequency oscillations *in both space and time*. Such PDE models arise in semiclassical modeling of quantum dynamics with band-crossings, and other highly oscillatory waves. Our first main idea is to use the nonlinear geometric optics ansatz, which builds the oscillatory phase into an independent variable. We then choose suitable initial data, based on the Chapman-Enskog expansion, for the new model. For a scalar model, we prove that so constructed model will have certain smoothness, and consequently, for a first order approximation scheme we prove uniform error estimates independent of the (possibly small) wave length. The method is extended to systems arising

from a semiclassical model for surface hopping, a non-adiabatic quantum dynamic phenomenon. Numerous numerical examples demonstrate that the method has the desired properties.

#### 4.1.6. High-order Hamiltonian splitting for Vlasov-Poisson equations

In [5], we consider the Vlasov-Poisson equation in a Hamiltonian framework and derive new time splitting methods based on the decomposition of the Hamiltonian functional between the kinetic and electric energy. Assuming smoothness of the solutions, we study the order conditions of such methods. It appears that these conditions are of Runge-Kutta-Nyström type. In the one dimensional case, the order conditions can be further simplified, and efficient methods of order 6 with a reduced number of stages can be constructed. In the general case, high-order methods can also be constructed using explicit computations of commutators. Numerical results are performed and show the benefit of using high-order splitting schemes in that context. Complete and self-contained proofs of convergence results and rigorous error estimates are also given.

# 4.1.7. A particle micro-macro decomposition based numerical scheme for collisional kinetic equations in the diffusion scaling

In [29], we derive particle schemes, based on micro-macro decomposition, for linear kinetic equations in the diffusion limit. Due to the particle approximation of the micro part, a splitting between the transport and the collision part has to be performed, and the stiffness of both these two parts prevent from uniform stability. To overcome this difficulty, the micro-macro system is reformulated into a continuous PDE whose coefficients are no longer stiff, and depend on the time step  $\Delta t$  in a consistent way. This non-stiff reformulation of the micro-macro system allows the use of standard particle approximations for the transport part, and extends a previous work where a particle approximation has been applied using a micro-macro decomposition on kinetic equations in the fluid scaling. Beyond the so-called asymptotic-preserving property which is satisfied by our schemes, they significantly reduce the inherent noise of traditional particle methods, and they have a computational cost which decreases as the system approaches the diffusion limit.

## 4.1.8. Uniformly accurate forward semi-Lagrangian methods for highly oscillatory Vlasov-Poisson equation

This work [13] is devoted to the numerical simulation of a Vlasov-Poisson equation modeling charged particles in a beam submitted to a highly oscillatory external electric field. A numerical scheme is constructed for this model. This scheme is uniformly accurate with respect to the size of the fast time oscillations of the solution, which means that no time step refinement is required to simulate the problem. The scheme combines the forward semi-Lagrangian method with a class of Uniformly Accurate (UA) time integrators to solve the characteristics. These UA time integrators are derived by means of a two-scale formulation of the characteristics, with the introduction of an additional periodic variable. Numerical experiments are done to show the efficiency of the proposed methods compared to conventional approaches.

# 4.1.9. Uniformly accurate multiscale time integrators for second order oscillatory differential equations with large initial data

In [23], we apply the modulated Fourier expansion to a class of second order differential equations which consists of an oscillatory linear part and a nonoscillatory nonlinear part, with the total energy of the system possibly unbounded when the oscillation frequency grows. We comment on the difference between this model problem and the classical energy bounded oscillatory equations. Based on the expansion, we propose the multiscale time integrators to solve the ODEs under two cases: the nonlinearity is a polynomial or the frequencies in the linear part are integer multiples of a single generic frequency. The proposed schemes are explicit and efficient. The schemes have been shown from both theoretical and numerical sides to converge with a uniform second order rate for all frequencies. Comparisons with popular exponential integrators in the literature are done.

# 4.1.10. Unconditional and optimal H<sup>2</sup>-error estimates of two linear and conservative finite difference schemes for the Klein-Gordon-Schrödinger equation in high dimensions

In [21], we focus on the optimal error bounds of two finite difference schemes for solving the *d*-dimensional (d = 2, 3) nonlinear Klein-Gordon-Schrödinger (KGS) equations. The proposed finite difference schemes not only conserve the mass and energy in the discrete level but also are efficient in practical computation because only two linear systems need to be solved at each time step. Besides the standard energy method, an induction argument as well as a 'lifting' technique are introduced to establish rigorously the optimal  $H^2$ -error estimates without any restrictions on the grid ratios, while the previous works either are not rigorous enough or often require certain restriction on the grid ratios. The convergence rates of the proposed schemes are proved to beat  $O(h^2 + \tau^2)$  with mesh size h and time step  $\tau$  in the discrete  $H^2$ -norm. The analysis method can be directly extended to other linear finite difference schemes for solving the KGS equations in high dimensions. Numerical results are reported to confirm the theoretical analysis for the proposed finite difference schemes

# 4.1.11. A combination of multiscale time integrator and two-scale formulation for the nonlinear Schrödinger equation with wave operator

In [22], we consider the nonlinear Schrödinger equation with wave operator (NLSW), which contains a dimensionless parameter  $0 < \varepsilon \leq 1$ . As  $0 < \varepsilon \ll 1$ , the solution of the NLSW propagates fast waves in time with wavelength  $O(\varepsilon^2)$  and the problem becomes highly oscillatory in time. The oscillations come from two parts. One part is from the equation and another part is from the initial data. For the ill-prepared initial data case as described in Bao and Cai (2014) which brings inconsistency in the limit regime, standard numerical methods have strong convergence order reduction in time when becomes small. We review two existing methods to solve the NLSW: an exponential integrator and a two-scale method. We comment on their order reduction issues. Then we derive a multiscale decomposition two-scale method for solving the NLSW by first performing a multiscale decomposition on the NLSW which decomposes it into a well-behaved part and an energy-unbounded part, and then applying an exponential integrator for the well-behaved part and a two-scale approach for the energy-unbounded part. Numerical experiments are conducted to test the proposed method which shows uniform second order accuracy without significant order reduction for all  $0 < \varepsilon \leq 1$ . Comparisons are made with the existing methods.

# 4.1.12. Uniformly accurate numerical schemes for the nonlinear Dirac equation in the nonrelativistic limit regime

In [18], we apply the two-scale formulation approach to propose uniformly accurate (UA) schemes for solving the nonlinear Dirac equation in the nonrelativistic limit regime. The nonlinear Dirac equation involves two small scales  $\varepsilon$  and  $\varepsilon^2$  with epsilon  $\rightarrow 0$  in the nonrelativistic limit regime. The small parameter causes high oscillations in time which brings severe numerical burden for classical numerical methods. We transform our original problem as a two-scale formulation and present a general strategy to tackle a class of highly oscillatory problems involving the two small scales  $\varepsilon$  and  $\varepsilon^2$ . Suitable initial data for the two-scale formulation is derived to bound the time derivatives of the augmented solution. Numerical schemes with uniform (with respect to  $\varepsilon \in (0; 1]$ ) spectral accuracy in space and uniform first order or second order accuracy in time are proposed. Numerical experiments are done to confirm the UA property.

#### 4.1.13. A formal series approach to the center manifold theorem

In [6], we consider near-equilibrium systems of ordinary differential equations with explicit separation of the slow and stable manifolds. Formal B-series like those previously used to analyze highly-oscillatory systems or to construct modified equations are employed here to construct expansions of the change of variables, the center invariant manifold and the reduced model. The new approach may be seen as a process of reduction to a normal form, with the main advantage, as compared to the standard view conveyed by the celebrated center manifold theorem, that it is possible to recover the complete solution at any time through an explicit change of variables.

# 4.1.14. Convergence of multi-revolution composition time-splitting methods for highly oscillatory differential equations of Schrödinger type

In [8], the convergence behaviour of multi-revolution composition methods combined with time-splitting methods is analysed for highly oscillatory linear differential equations of Schrödinger type. Numerical experiments illustrate and complement the theoretical investigations.

# 4.1.15. Highly-oscillatory evolution equations with multiple frequencies: averaging and numerics

In [7], we are concerned with the application of the recently introduced multi-revolution composition methods, on the one hand, and two-scale methods, on the other hand, to a class of highly-oscillatory evolution equations with multiple frequencies. The main idea relies on a well-balanced reformulation of the problem as an equivalent mono-frequency equation which allows for the use of the two aforementioned techniques.

#### 4.1.16. Optimality and resonances in a class of compact finite difference schemes of high order

In [25], we revisit the old problem of compact finite difference approximations of the homogeneous Dirichlet problem in dimension 1. We design a large and natural set of schemes of arbitrary high order, and we equip this set with an algebraic structure. We give some general criteria of convergence and we apply them to obtain two new results. On the one hand, we use Padé approximant theory to construct, for each given order of consistency, the most efficient schemes and we prove their convergence. On the other hand, we use diophantine approximation theory to prove that almost all of these schemes are convergent at the same rate as the consistency order, up to some logarithmic correction.

## 4.2. mathematical analysis of multiscale partial differential equations

### 4.2.1. Collision of almost parallel vortex filaments

In [3], we investigate the occurrence of collisions in the evolution of vortex filaments through a system introduced by Klein, Majda and Damodaran and Zakharov. We first establish rigorously the existence of a pair of almost parallel vortex filaments, with opposite circulation, colliding at some point in finite time. The collision mechanism is based on the one of the self-similar solutions of the model, described in a previous work. In the second part of this paper we extend this construction to the case of an arbitrary number of filaments, with polygonial symmetry, that are perturbations of a configuration of parallel vortex filaments forming a polygon, with or without its center, rotating with constant angular velocity.

#### 4.2.2. Free vibrations of axisymmetric shells: parabolic and elliptic cases

In [9], approximate eigenpairs (quasimodes) of axisymmetric thin elastic domains with laterally clamped boundary conditions (Lamé system) are determined by an asymptotic analysis as the thickness  $(2\varepsilon)$  tends to zero. The departing point is the Koiter shell model that we reduce by asymptotic analysis to a scalar model that depends on two parameters: the angular frequency k and the half-thickness  $\varepsilon$ . Optimizing k for each chosen  $\varepsilon$ , we find power laws for k in function of  $\varepsilon$  that provide the smallest eigenvalues of the scalar reductions. Corresponding eigenpairs generate quasimodes for the 3D Lamé system by means of several reconstruction operators, including boundary layer terms. Numerical experiments demonstrate that in many cases the constructed eigenpair corresponds to the first eigenpair of the Lamé system. Geometrical conditions are necessary to this approach: The Gaussian curvature has to be nonnegative and the azimuthal curvature has to dominate the meridian curvature in any point of the midsurface. In this case, the first eigenvector admits progressively larger oscillation in the angular variable as  $\varepsilon$  tends to 0. Its angular frequency exhibits a power law relation of the form  $k = \gamma \varepsilon^{\beta}$  with  $\beta = \frac{1}{4}$  in the parabolic case (cylinders and trimmed cones), and the various  $\beta$ 's  $(\frac{2}{5}, \frac{3}{7}$  and  $\frac{1}{3}$  in the elliptic case. For these cases where the mathematical analysis is applicable, numerical examples that illustrate the theoretical results are presented.

# 4.2.3. High frequency oscillations of first eigenmodes in axisymmetric shells as the thickness tends to zero

In [24], the lowest eigenmode of thin axisymmetric shells is investigated for two physical models (acoustics and elasticity) as the shell thickness  $(2\varepsilon)$  tends to zero. Using a novel asymptotic expansion we determine the behavior of the eigenvalue  $\lambda(\varepsilon)$  and the eigenvector angular frequency  $k(\varepsilon)$  for shells with Dirichlet boundary conditions along the lateral boundary, and natural boundary conditions on the other parts. First, the scalar Laplace operator for acoustics is addressed, for which  $k(\varepsilon)$  is always zero. In contrast to it, for the Lamé system of linear elasticity several different types of shells are defined, characterized by their geometry, for which  $k(\varepsilon)$  tends to infinity as  $\varepsilon$  tends to zero. For two families of shells: cylinders and elliptical barrels we explicitly provide  $\lambda(\varepsilon)$  and  $k(\varepsilon)$  and demonstrate by numerical examples the different behavior as  $\varepsilon$  tends to zero.

#### 4.2.4. Semiclassical Sobolev constants for the electro-magnetic Robin Laplacian

This paper [15] is devoted to the asymptotic analysis of the optimal Sobolev constants in the semiclassical limit and in any dimension. We combine semiclassical arguments and concentration-compactness estimates to tackle the case when an electromagnetic field is added as well as a smooth boundary carrying a Robin condition. As a byproduct of the semiclassical strategy, we also get exponentially weighted localization estimates of the minimizers.

#### 4.2.5. On the MIT Bag Model in the Non-relativistic Limit

This paper [2] is devoted to the spectral investigation of the MIT bag model, that is, the Dirac operator on a smooth and bounded domain of  $\mathbb{R}^3$  with certain boundary conditions. When the mass m goes to  $\pm \infty$ , we provide spectral asymptotic results.

# 4.2.6. Dimension reduction for dipolar Bose-Einstein condensates in the strong interaction regime

In [4], we study dimension reduction for the three-dimensional Gross-Pitaevskii equation with a long-range and anisotropic dipole-dipole interaction modeling dipolar Bose-Einstein condensation in a strong interaction regime. The cases of disk shaped condensates (confinement from dimension three to dimension two) and cigar shaped condensates (confinement to dimension one) are analyzed. In both cases, the analysis combines averaging tools and semiclassical techniques. Asymptotic models are derived, with rates of convergence in terms of two small dimensionless parameters characterizing the strength of the confinement and the strength of the interaction between atoms.

#### 4.2.7. Nonlinear stability criteria for the HMF Model

In [17], we study the nonlinear stability of a large class of inhomogeneous steady state solutions to the Hamiltonian Mean Field (HMF) model. Under a specific criterion, we prove the nonlinear stability of steady states which are decreasing functions of the microscopic energy. To achieve this task, we extend to this context the strategy based on generalized rearrangement techniques which was developed recently for the gravitational Vlasov-Poisson equation. Explicit stability inequalities are established and our analysis is able to treat non compactly supported steady states to HMF, which are physically relevant in this context but induces additional difficulties, compared to the Vlasov-Poisson system.

#### 4.2.8. Strong confinement limit for the nonlinear Schrödinger equation constrained on a curve

This paper [20] is devoted to the cubic nonlinear Schrödinger equation in a two dimensional waveguide with shrinking cross section. For a Cauchy data living essentially on the first mode of the transverse Laplacian, we provide a tensorial approximation of the solution in this limit, with an estimate of the approximation error, and derive a limiting nonlinear Schrödinger equation in dimension one.

#### 4.2.9. Stable ground states for the HMF Poisson Model

In [36], we prove the nonlinear orbital stability of a large class of steady states solutions to the Hamiltonian Mean Field (HMF) system with a Poisson interaction potential. These steady states are obtained as minimizers of an energy functional under one, two or infinitely many constraints. The singularity of the Poisson potential prevents from a direct run of the general strategy which was based on generalized rearrangement techniques, and which has been recently extended to the case of the usual (smooth) cosine potential. Our strategy is rather based on variational techniques. However, due to the boundedness of the space domain, our variational problems do not enjoy the usual scaling invariances which are, in general, very important in the analysis of variational problems. To replace these scaling arguments, we introduce new transformations which, although specific to our context, remain somehow in the same spirit of rearrangements tools introduced in the references above. In particular, these transformations allow for the incorporation of an arbitrary number of constraints, and yield a stability result for a large class of steady states.

#### 4.2.10. The quantum Liouville-BGK equation and the moment problem

This work [19] is devoted to the analysis of the quantum Liouville-BGK equation. This equation arises in the work of Degond and Ringhofer on the derivation of quantum hydrodynamical models from first principles. Their theory consists in transposing to the quantum setting the closure strategy by entropy minimization used for kinetic equations. The starting point is the quantum Liouville-BGK equation, where the collision term is defined via a so-called quantum local equilibrium, defined as a minimizer of the quantum free energy under a local density constraint. We then address three related problems: we prove new results about the regularity of these quantum equilibria; we prove that the quantum Liouville-BGK equation admits a classical solution; and we investigate the long-time behavior of the solutions. The core of the proofs is based on a fine analysis of the properties of the minimizers of the free energy.

#### 4.2.11. Averaging of nonlinear Schrödinger equations with strong magnetic confinement

In [16], we consider the dynamics of nonlinear Schrödinger equations with strong constant magnetic fields. In an asymptotic scaling limit the system exhibits a purely magnetic confinement, based on the spectral properties of the Landau Hamiltonian. Using an averaging technique we derive an associated effective description via an averaged model of nonlinear Schrödinger type. In a special case this also yields a derivation of the LLL equation.

## 4.3. mathematical analysis of stochastic partial differential equations

### 4.3.1. Large deviations for the dynamic $\Phi_d^{2n}$ model

In [27], we are dealing with the validity of a large deviation principle for a class of reaction-diffusion equations with polynomial non-linearity, perturbed by a Gaussian random forcing. We are here interested in the regime where both the strength of the noise and its correlation are vanishing, on a length scale  $\rho$  and  $\delta(\rho)$ , respectively, with  $0 < \rho, \delta(\rho) \ll 1$ . We prove that, under the assumption that  $\rho$  and  $\delta(\rho)$  satisfy a suitable scaling limit, a large deviation principle holds in the space of continuous trajectories with values both in the space of square-integrable functions and in Sobolev spaces of negative exponent. Our result is valid, without any restriction on the degree of the polynomial nor on the space dimension.

#### 4.3.2. Solution to the stochastic Schrödinger equation on the full space

In [33], we show how the methods recently applied by Debussche and Weber to solve the stochastic nonlinear Schrödinger equation on  $\mathbb{T}^2$  can be enhanced to yield solutions on  $\mathbb{R}^2$  if the non-linearity is weak enough. We prove that the solutions remains localized on compact time intervals which allows us to apply energy methods on the full space.

# 4.3.3. A law of large numbers in the supremum norm for a multiscale stochastic spatial gene network

In [34], we study the asymptotic behavior of multiscale stochastic spatial gene networks. Multiscaling takes into account the difference of abundance between molecules, and captures the dynamic of rare species at a mesoscopic level. We introduce an assumption of spatial correlations for reactions involving rare species and a new law of large numbers is obtained. According to the scales, the whole system splits into two parts with different but coupled dynamics. The high scale component converges to the usual spatial model which is the solution of a partial differential equation, whereas, the low scale component converges to the usual homogeneous model which is the solution of an ordinary differential equation. Comparisons are made in the supremum norm.

### 4.3.4. Long time behavior of Gross-Pitaevskii equation at positive temperature

In [32], the stochastic Gross-Pitaevskii equation is used as a model to describe Bose-Einstein condensation at positive temperature. The equation is a complex Ginzburg Landau equation with a trapping potential and an additive space-time white noise. Two important questions for this system are the global existence of solutions in the support of the Gibbs measure, and the convergence of those solutions to the equilibrium for large time. In this paper, we give a proof of these two results in one space dimension. In order to prove the convergence to equilibrium, we use the associated purely dissipative equation as an auxiliary equation, for which the convergence may be obtained using standard techniques.

# 4.3.5. An integral inequality for the invariant measure of a stochastic reaction–diffusion equation

In [14], we consider a reaction-diffusion equation perturbed by noise (not necessarily white). We prove an integral inequality for the invariant measure  $\nu$  of a stochastic reaction-diffusion equation. Then we discuss some consequences as an integration by parts formula which extends to  $\nu$  a basic identity of the Malliavin Calculus. Finally, we prove the existence of a surface measure for a ball and a half-space of  $\mathcal{H}$ .

# 4.3.6. Kolmogorov equations and weak order analysis for SPDES with nonlinear diffusion coefficient

In [26], we provide new regularity results for the solutions of the Kolmogorov equation associated to a SPDE with nonlinear diffusion coefficients and a Burgers type nonlinearity. This generalizes previous results in the simpler cases of additive or affine noise. The basic tool is a discrete version of a two sided stochastic integral which allows a new formulation for the derivatives of these solutions. We show that this can be used to generalize the weak order analysis performed by Debussche in 2011. The tools we develop are very general and can be used to study many other examples of applications.

#### 4.3.7. Approximation-diffusion in stochastically forced kinetic equations

In [35], we derive the hydrodynamic limit of a kinetic equation where the interactions in velocity are modelled by a linear operator (Fokker-Planck or Linear Boltzmann) and the force in the Vlasov term is a stochastic process with high amplitude and short-range correlation. In the scales and the regime we consider, the hydrodynamic equation is a scalar second-order stochastic partial differential equation. Compared to the deterministic case, we also observe a phenomenon of enhanced diffusion.

# 5. Partnerships and Cooperations

# 5.1. Regional Initiatives

- A. Crestetto is member of the project "Pari Scientifique Régional Exprodil".
- M. Lemou is member of the project "Défis" of the University of Rennes 1, leader Nicolas Seguin.

# 5.2. National Initiatives

#### 5.2.1. ANR MOONRISE: 2015-2019

Participants: Francois Castella, Philippe Chartier, Nicolas Crouseilles, Mohammed Lemou, Florian Mehats.

The project *Moonrise* submitted by Florian Méhats has been funded by the ANR for 4 years, for the period 2015-2019. This project aims at exploring modeling, mathematical and numerical issues originating from the presence of high-oscillations in nonlinear PDEs from the physics of nanotechnologies (quantum transport) and from the physics of plasmas (magnetized transport in tokamaks). The partners of the project are the IRMAR (Rennes), the IMT (Toulouse) and the CEA Cadarache. In the IPSO team, François Castella, Philippe Chartier, Nicolas Crouseilles and Mohammed Lemou are members of the project Moonrise.

#### Postdocs

- Loïc Le Treust has been hired as a Postdoc, under the supervision of Philippe Chartier and Florian Méhats. His contract started in september 2015 and ended in august 2016. Loïc Le Treust is now assistant professor at the university of Marseille.
- Xiaofei Zhao has been hired as a Postdoc from september 2015 to september 2016 under the supervision of Florian Méhats.

#### 5.2.2. ANR MFG: 2016-2020

#### Participant: Arnaud Debussche.

Mean Field Games (MFG) theory is a new and challenging mathematical topic which analyzes the dynamics of a very large number of interacting rational agents. Introduced ten years ago, the MFG models have been used in many areas such as, e.g., economics (heterogeneous agent models, growth modeling,...), finance (formation of volatility, models of bank runs,...), social sciences (crowd models, models of segregation) and engineering (data networks, energy systems...). Their importance comes from the fact that they are the simplest ("stochastic control"-type) models taking into account interactions between rational agents (thus getting beyond optimization), yet without entering into the issues of strategic interactions. MFG theory lies at the intersection of mean field theories (it studies systems with a very large number of agents), game theory, optimal control and stochastic analysis (the agents optimize a payoff in a possibly noisy setting), calculus of variations (MFG equilibria may arise as minima of suitable functionals) and partial differential equations (PDE): In the simplest cases, the value of each agent is found by solving a backward Hamilton-Jacobi equation whereas the distribution of the agents' states evolves according to a forward Fokker-Planck equation. The "Master" equation (stated in the space of probability measures) subsumes the individual and collective behaviors. Finally, modeling, numerical analysis and scientific computing are crucial for the applications. French mathematicians play a world-leading role in the research on MFG: The terminology itself comes from a series of pioneering works by J.-M. Lasry and P.-L. Lions who introduced most of the key ideas for the mathematical analysis of MFG; the last conference on MFG was held last June in Paris and organized by Y. Achdou, P. Cardaliaguet and J.-M. Lasry. As testifies the proposal, the number of researchers working on MFG in France (and also abroad) is extremely fast-growing, not only because the theoretical aspects are exciting and challenging, but also because MFG models find more and more applications. The aim of the project is to better coordinate the French mathematical research on MFG and to achieve significant progress in the theory and its applications.

The partners of the project are the CEREMADE laboratory (Paris Dauphine), the IRMAR laboratory (Rennes I), the university of Nice and of Tours.

#### 5.2.3. ANR ACHYLLES: 2014-2018

Participant: Anais Crestetto.

The ACHYLLES project focuses on Long-Time Asymptotic-Preserving (LTAP) numerical schemes for hyperbolic systems of conservation laws supplemented by potentially stiff source terms. It ambitions to perform a breakthrough in the understanding and efficiency of LTAP scheme.

The partners are IMB (Bordeaux), LMV (Versailles) and LMJL (Nantes).

### 5.2.4. IPL FRATRES

IPSO is associated to IPL FRATRES which started in june 2015. The aim of this project is to organize Inria teams activities which develop mathematical and numerical tools in magnetically confined nuclear fusion. The ambition is to prepare the next generation of numerical modeling methodologies able to use in an optimal way the processing capabilities of modern massively parallel architectures. This objective requires close collaboration between a) applied mathematicians and physicists that develop and study mathematical models of PDE; b) numerical analysts developing approximation schemes; c) specialists of algorithmics proposing solvers and libraries using the many levels of parallelism offered by the modern architecture and d) computer scientists. The project road map ambitions to contribute in close connection with National and European initiatives devoted to nuclear Fusion to the improvement and design of numerical simulation technologies applied to plasma physics and in particular to the ITER project for magnetic confinement fusion.

#### Postdoc

• Xiaofei Zhao has been hired as a Postdoc, under the supervision of Nicolas Crouseilles and Sever Hirstoaga (Inria-Nancy). His contract started in october 2016 and ended in september 2017.

# **5.3. European Initiatives**

#### 5.3.1. Collaborations in European Programs, Except FP7 & H2020

Program: EUROFusion Enabling Research

Project acronym: WPENR

Project title: Verification and development of new algorithms for gyrokinetic codes

Duration: january 2015 - december 2017.

Coordinator: E. Sonnendrücker (Max-Planck IPP, Germany)

Other partners: IPP, EPFL, CEA-Cadarache, university of Strasbourg, Toulouse, Marseille, Paris 6.

Abstract: Gyrokinetic codes play a major role in understanding the development and saturation of micro-turbulence in a magnetic fusion plasma and its influence on energy confinement time. The first aim of this proposal is to assess the reliability of gyrokinetic codes by extensive verification and benchmarking. All the major european gyrokinetic codes are involved in the proposal and this will enable them to define comparison elements, which ultimately will also facilitate the cross-validation of new physics. On the other hand we will develop new algorithms for extending the physics capabilities or the computational efficiency of different gyrokinetic codes. Finally we will also perform a prospective investigation of models and numerical methods that could help in the future to address physics where kinetic effects might play an important role but that cannot be handled with today's gyrokinetic codes, like L-H (low to high confinement) transition, edge physics or MHD time scales simulations.

## **5.4. International Initiatives**

#### 5.4.1. Informal International Partners

Members of the IPSO team have several international collaborations

- the group of S. Jin (university of Wisconsin, US).
- the group of W. Bao (university of Singapore).
- G. Vimart (university of Geneva, Swizerland).
- the group of A. Ostermann (university of Innsbrück).
- the SNS of Pisa (G. Da Prato).
- several US universities: Maryland (S. Cerrai), Chicago (C. Sparber), Colorado (O. Pinaud), ...
- • • •

### 5.4.2. Participation in Other International Programs

• A. Crestetto is involved in the project PHC PROCOPE "Hétérogénéités Fortes dans les Modèles d'Ecoulement Fluide".

# 5.5. International Research Visitors

## 5.5.1. Visits to International Teams

- P. Chartier was invited by Fernando Casas, University of Castellon, Spain, July 6-9 2017.
- A. Crestetto was invited by Christian Klingenberg, Institute of Mathematics, Würzburg University, July 10-14 2017.
- M. Lemou was invited by Shi Jin, Jiao Tong university Shanghai, China, July 5-15 2017.
- M. Lemou was invited by Hao Wu, Tsinghua university Beijing, China, July 15-20 2017.

# 6. Dissemination

# 6.1. Promoting Scientific Activities

### 6.1.1. Scientific Events Organisation

6.1.1.1. Member of the Organizing Committees

- F. Castella organized the workshop "Multiscale numerical methods", Saint-Malo, december 13-15 2017. [15 participants]
- N. Crouseilles co-organized the weekly seminar "Mathematic and applications", ENS Rennes.
- N. Crouseilles co-organized the Nantes-Rennes meeting at the university of Nantes, january 19 2017.
   [50 participants]
- N. Crouseilles organized the IPL FRATRES meeting at Inria Rennes, november 27-28 2017. [30 participants]
- N. Crouseilles and M. Lemou co-organized (with C. Cheverry and K. Pravda-Starov) the international workshop "Analysis of transport equations: Vlasov and related models", Rennes, may 16-19 2017. [50 participants]
- E. Faou organizes a work-group on "Mathematics of deep learning" (Inria, IRMAR and Technicolor), Rennes, France.
- E. Faou organized the mini-sympositum "Methods for the nonlinear Schrödinger equations, solitary waves and discrete patterns", with T. Matsuo (University of Tokyo), at the Scicade international conference, Bath (UK), september 11-15 2017.
- E. Faou co-organized the conference ANSIVAL on the occation of the 60th birthday of M. Dauge (IRMAR), Rennes, France, february 8-10 2017. [70 participants]

### 6.1.2. Journal

#### 6.1.2.1. Member of the Editorial Boards

- P. Chartier is associate editor of M2AN (2007-).
  - A. Debussche is editor in chief of Stochastic Partial Differential Equations: analysis and computations (2013-).
  - A. Debussche is associate editor of Differential and Integral Equations (2002-).
  - A. Debussche is associate editor of Potential Analysis (2011-).
  - A. Debussche is associate editor of ESAIM:PROC (2012-).
  - A. Debussche is associate editor of Journal of Evolution Equation (2014-).
  - A. Debussche is associate editor of Applied Mathematics & Optimization (2014-).
  - A. Debussche is associate editor of SIAM JUQ (2017-).
  - A. Debussche is member of the editorial board of the collection "Mathématiques & Applications de la SMAI", Springer.

#### 6.1.2.2. Reviewer - Reviewing Activities

The members of the IPSO team are reviewers for almost all the journals in which they publish.

#### 6.1.3. Invited Talks

National conferences

- A. Crestetto gave talk at Congrès SMAI 2017, La Tremblade, june 5-9 2017.
- The members of the team gave several seminars in french institutions (Marseille, Cergy, Toulouse, Nancy, CEA Cadarache, ...).

International conferences

- P. Chartier was invited speaker at FoCM 2017, workshop on "Geometric Integration and Computational Mechanics" (july 10-12), Barcelona, Spain.
- P. Chartier was invited speaker at "Mathematical and Computational Methods for Quantum and Kinetic Problems", (june 12-14), Beijing, China.
- P. Chartier was invited speaker at "Kinema 2017: Numerical Modelling of Kinetic Magnetized Plasmas", spring school, Institut d'études scientifiques de Cargèse, (april 3-7), Cargèse, France.
- A. Crestetto gave talk at the workshop "NumKin 2017" (october 23-27), IPP Garching, Germany.
- A. Crestetto gave a talk at the Oberseminar (july 11), Würzburg, Germany.
- N. Crouseilles gave a talk at the workshop on "Mathematical and Computational methods for Quantum and Kinetic Problems, (june 12-14), Beijing, China.
- A. Debussche gave a talk at "Stochastic PDEs: Analysis and Computation", (march 27-31), Warwick, England.
- A. Debussche gave a talk at "Probabilistic Perspectives in Nonlinear PDEs", (june 5-9), International Centre for Mathematical Sciences (ICMS), Edinburgh, Scotland.
- A. Debussche gave a talk at FoCM2017, workshop on "Stochastic Computation", (july 10-12), Barcelona, Spain.
- A. Debussche gave a talk at "2017 Fields Medal Symposium", (october 16-19), Toronto, Canada.
- E. Faou gave a talk at the mini-symposium "Modelling, theory and approximation of nonlinear waves", Scicade international conference, (september 11-15), University of Bath, UK.
- E. Faou gave a talk at the conference "Asymptotic analysis of evolution equations", (july 3-7), CIRM, Marseille.
- E. Faou gave a talk at the seminar at the University of Cambridge, (june 2017), UK.
- E. Faou gave a talk at the workshop "Mathematical questions in wave turbulence theory", (may 15-19), San Jose, California.
- E. Faou gave a talk at the workshop "Modern Numerical Methods for Quantum Mechanics", (march 20-22) Polish Academy of sciences, Warsaw, Poland.
- M. Lemou gave a talk at International workshop "Geometric Transport Equations in General Relativity", (february 20-24), ESI, Vienna, Austria.
- M. Lemou gave a course in the summer school "Applied and Stochastic Analysis for Partial Differential Equations". Institute of Natural Science, Shanghai Jiao Tong University, (july 12-22), Shanghai, China.
- M. Lemou gave a talk at WPI workshop on "Quantum Dynamics and Uncertainty Quantification", (june 20-25), Vienna, Austria,
- M. Lemou gave a talk at the workshop on "Mathematical and Computational methods for Quantum and Kinetic Problems, (june 12-14), Beijing, China.
- M. Lemou gave a course at "Kinema 2017: Numerical Modelling of Kinetic Magnetized Plasmas", spring school, Institut d'études scientifiques de Cargèse, (april 3-7), Cargèse, France.
- M. Lemou gave a talk at the workshop on "Kinetic Theory and Fluid Mechanics : theoretical and computational aspects", (november 6-10), Toulouse, France.

- F. Méhats gave a talk at SIAM Conference on "Analysis of Partial Differential Equations", (december 10-12), Baltimore, USA.
- F. Méhats gave a talk at the workshop on "Mathematical and Computational methods for Quantum and Kinetic Problems, (june 12-14), Beijing, China.
- F. Méhats gave a talk at the conference "Advances in Mathematics for Technology", (october 9-11) Catania, Italy.

#### 6.1.4. Scientific Expertise

- P. Chartier is member of the promotion committees DR1 and DR0 Inria.
- N. Crouseilles was reviewer for ANR project.
- A. Debussche participated to the report projet EUR Centre Henri Lebesgue.
- A. Debussche was reviewer for ERC projects, "Philip Leverhulme" (GB) fundation, Austrian Science Fundation.
- A. Debussche was president of the visiting committee HCERES of the Centre d'Analyses et de Mathématique Sociales (Paris, EHESS).

#### 6.1.5. Research Administration

- P. Chartier is scientific vice-deputy of the Inria Rennes center.
- P. Chartier is member of the Inria evaluation committee.
- P. Chartier is member of the Inria Scientific Committee (COSI).
- P. Chartier is member of the Bureau du Comité des Projets (BCP).
- A. Crestetto is member of the mathematic department council of the university of Nantes.
- A. Crestetto is member of the scientific council of "UFR Sciences et Techniques" of the university of Nantes.
- N. Crouseilles is member of the scientific council of ENS Rennes.
- N. Crouseilles is member of the mathematic laboratory (IRMAR) council.
- N. Crouseilles is member of the Fédération de Fusion council (University of Rennes representative).
- A. Debussche is member of the scientific council of the Fédération Denis Poisson.
- A. Debussche is member of the administrative council of ENS Paris-Saclay.
- A. Debussche if scientific vice-deputy and international relations of ENS Rennes.
- A. Debussche is vice-head ot the Centre Henri Lebesgue.
- A. Debussche is vice-head of the Lebesgue agency for Mathematic and Innovation.
- E. Faou is member of the scientific council of the Pôle Universitaire Léonard de Vinci.
- E. Faou is member of the CNU section 26.
- E. Faou is head of organization of the semester *scientific computing* sponsored by the Labex Lebesgue (2 international summer schools, 7 workshops and international conferences).
- M. Lemou is member of the scientific council of the Center Henri Lebesgue.
- M. Lemou is member of the scientific council of ENS Rennes.
- M. Lemou is the head of the numerical analysis team of IRMAR laboratory. [46 members].
- F. Méhats was head of the mathematic laboratory IRMAR (2015-2017). [250 members].

# 6.2. Teaching - Supervision - Juries

#### 6.2.1. Teaching

• Master : F. Castella, "Equations de transport et Phenomenes de Propagation", 48h, M1, university of Rennes 1, France.

- Master : F. Castella, "Analyse Numerique Generale", 48h, M1, university of Rennes 1, France.
- Master : P. Chartier, "Semi-lagrangian methods for Vlasov-Poisson equations", 18h, M2, university of Rennes 1, France.
- Master : A. Crestetto, "Méthodes numériques pour les fluides incompressibles", 64h, M2, university of Nantes, France.
- Master : A. Crestetto, "Compléments de modélisation pour l'agrégation", 26h, M2, university of Nantes, France.
- Master : N. Crouseilles, "Analyse numérique", 30h, M1, ENS Rennes, France.
- Master : A. Debussche, "Distribution et analyse fonctionnelle", 30h, M1, ENS Rennes, France.
- Master : A. Debussche, "Introduction aux EDP Stochastiques", 48h, M2, university of Rennes 1, France.
- Master : A. Debussche, "Compléments pour l'agrégation", 26h, M2, ENS Rennes, France.
- Master : M. Lemou, "Equations aux dérivées partielles elliptiques", 30h, M1, ENS de Rennes.
- Master : F. Méhats, "Equations hyperboliques", 30h, M2, university of Rennes 1, France.

#### 6.2.2. Supervision

- PhD : R. Horsin, Comportement en temps long d'équations de type Vlasov: Etudes mathématiques et numériques, university of Rennes 1, december 1st 2017, E. Faou and F. Rousset (university Paris Sud).
- PhD : V. Doli, Phénomènes de propagation de champignons parasites de plantes, par couplage de diffusion spatiale et de reproduction sexuée, december 23th 2017, F. Castella and F. Hamelin (IGEPP, Agrocampus).
- PhD in progress : M. Malo, Equations cinétiques non collisionnelles : stabilité, oscillations, september 2015, M. Lemou and F. Méhats.
- PhD in progress : J. Bernier, Mathematical and numerical analysis of nonlinear transport equations, (2016-), september 2016, N. Crouseilles and E. Faou.
- PhD in progress : M. Tusseau, Sur l'équation de Schrödinger non linéaire hautement oscillante avec potentiel aléatoire, september 2013, A. Debussche and F. Méhats.
- PhD in progress : M. Jugal Nguepedja Nankep, Modèles spatiaux stochastiques de systèmes multiéchelle de particules en interactions, september 2014, A. Debussche.
- PhD in progress : A. Rosello, Approximation-diffusion pour des équations cinétiques pour les modèles de type "spray", september 2017, A. Debussche.

#### 6.2.3. Juries

- A. Crestetto was member of the jury of the thesis of T. Blanc, "Etude mathématique de problèmes paraboliques fortement anisotropes", Marseille, december 4th 2017.
- N. Crouseilles was member of the jury of the thesis of A. Finot, "Analyse mathématique des modèles cinétiques en présence d'un champ magnétique intense", Marseille, january 26th 2017.
- N. Crouseilles was member of the jury of the thesis of T. Hardy, "Traitement des conditions aux limites spéculaires pour l'étude du transfert radiatif dans des matériaux à géométrie complexe", Nantes, january 31th 2017.
- E. Faou was member of the jury of the thesis of S. Dieckmann, "Dynamics of patterns in equivariant Hamiltonian partial differential equations", Bielefeld (Germany), april 2017.
- E. Faou was member of the jury of the thesis of P. Krämer, "Numerical integrators for Maxwell-Klein-Gordon and Maxwell-Dirac systems in highly to slowly oscillatory regimes", Karlsruhe (Germany), august 2017.

- M. Lemou was member of the jury of the thesis of R. Horsin Blanc, "Comportement en temps long d'équations de type Vlasov: Etudes mathématiques et numériques", Rennes, december 1st 2017.
- M. Lemou was member of the jury (reviewer) of the thesis of T. Blanc, "Etude mathématique de problèmes paraboliques fortement anisotropes", Marseille, december 4th 2017.
- F. Méhats was member of the jury of the thesis of A. Finot, "Analyse mathématique des modèles cinétiques en présence d'un champ magnétique intense", Marseille, january 26th 2017.

# 7. Bibliography

## **Publications of the year**

#### **Doctoral Dissertations and Habilitation Theses**

 R. HORSIN. On long time behavior of certain Vlasov equations: Mathematics and Numerics, Université de Rennes 1, December 2017, https://tel.archives-ouvertes.fr/tel-01670352.

#### **Articles in International Peer-Reviewed Journal**

- [2] N. ARRIZABALAGA, L. LE TREUST, N. RAYMOND. On the MIT Bag Model in the Non-relativistic Limit, in "Communications in Mathematical Physics", 2017, vol. 354, n<sup>o</sup> 2, p. 641-669 [DOI: 10.1007/s00220-017-2916-8], https://hal.archives-ouvertes.fr/hal-01343717.
- [3] V. BANICA, E. FAOU, E. MIOT. Collision of almost parallel vortex filaments, in "Communications on Pure and Applied Mathematics", 2017, vol. 70, n<sup>o</sup> 2, p. 378-405 [DOI : 10.1002/CPA.21637], https://hal.archivesouvertes.fr/hal-01170929.
- [4] W. BAO, L. LE TREUST, F. MÉHATS.Dimension reduction for dipolar Bose-Einstein condensates in the strong interaction regime, in "Kinetic and Related Models ", September 2017, vol. 10, n<sup>o</sup> 3, p. 553-571, https://arxiv. org/abs/1501.02177 [DOI: 10.3934/KRM.2017022], https://hal.archives-ouvertes.fr/hal-01101793.
- [5] F. CASAS, N. CROUSEILLES, E. FAOU, M. MEHRENBERGER.*High-order Hamiltonian splitting for Vlasov-Poisson equations*, in "Numerische Mathematik", 2017, vol. 135, n<sup>o</sup> 3, p. 769-801, https://arxiv.org/abs/1510. 01841 [DOI: 10.1007/s00211-016-0816-z], https://hal.inria.fr/hal-01206164.
- [6] F. CASTELLA, P. CHARTIER, J. SAUZEAU.A formal series approach to the center manifold theorem, in "Foundations of Computational Mathematics", 2017 [DOI : 10.1007/s10208-017-9371-Y], https://hal. inria.fr/hal-01279715.
- [7] P. CHARTIER, M. LEMOU, F. MÉHATS.*Highly-oscillatory evolution equations with multiple frequencies: averaging and numerics*, in "Numerische Mathematik", December 2017, vol. 136, n<sup>o</sup> 4, p. 907-939 [DOI: 10.1007/s00211-016-0864-4], https://hal.inria.fr/hal-01281950.
- [8] P. CHARTIER, F. MÉHATS, M. THALHAMMER, Y. ZHANG. Convergence of multi-revolution composition time-splitting methods for highly oscillatory differential equations of Schrödinger type, in "ESAIM: Mathematical Modelling and Numerical Analysis", September 2017, vol. 51, n<sup>o</sup> 5, p. 1859 - 1882 [DOI: 10.1051/M2AN/2017010], https://hal.archives-ouvertes.fr/hal-01636323.

- [9] M. CHAUSSADE-BEAUDOUIN, M. DAUGE, E. FAOU, Z. YOSIBASH. Free Vibrations of Axisymmetric Shells: Parabolic and Elliptic cases, in "Asymptotic Analysis", 2017, vol. 104, n<sup>o</sup> 2, p. 1-47, https://arxiv.org/abs/ 1602.00850 [DOI: 10.3233/ASY-171426], https://hal.archives-ouvertes.fr/hal-01264125.
- [10] N. CROUSEILLES, G. DIMARCO, M. LEMOU.Asymptotic preserving and time diminishing schemes for rarefied gas dynamic, in "Kinetic and Related Models ", 2017, vol. 10, p. 643-668, https://hal.inria.fr/hal-01392412.
- [11] N. CROUSEILLES, S. A. HIRSTOAGA, X. ZHAO. Multiscale Particle-in-Cell methods and comparisons for the long-time two-dimensional Vlasov-Poisson equation with strong magnetic field, in "Computer Physics Communications", October 2017, vol. 222, p. 136-151, https://hal.archives-ouvertes.fr/hal-01496854.
- [12] N. CROUSEILLES, S. JIN, M. LEMOU.Nonlinear Geometric Optics method based multi-scale numerical schemes for highly-oscillatory transport equations, in "Mathematical Models and Methods in Applied Sciences", 2017, vol. 27, n<sup>o</sup> 11, p. 2031-2070 [DOI : 10.1142/S0218202517500385], https://hal.archivesouvertes.fr/hal-01323721.
- [13] N. CROUSEILLES, M. LEMOU, F. MÉHATS, X. ZHAO. Uniformly accurate forward semi-Lagrangian methods for highly oscillatory Vlasov-Poisson equations, in "Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal", 2017, vol. 15, n<sup>o</sup> 2, p. 723-744 [DOI: 10.1137/16M1059497], https://hal.inria.fr/ hal-01286947.
- [14] G. DA PRATO, A. DEBUSSCHE.An integral inequality for the invariant measure of a stochastic reaction-diffusion equation, in "Journal of Evolution Equations", 2017, vol. 17, n<sup>o</sup> 1, p. 197-214, https://arxiv. org/abs/1511.07133 [DOI: 10.1007/s00028-016-0349-z], https://hal.archives-ouvertes.fr/hal-01235038.
- [15] S. FOURNAIS, L. LE TREUST, N. RAYMOND, J. VAN SCHAFTINGEN. Semiclassical Sobolev constants for the electro-magnetic Robin Laplacian, in "Journal of the Mathematical Society of Japan", 2017, vol. 69, n<sup>0</sup> 4, p. 1667-1714, https://arxiv.org/abs/1603.02810 [DOI: 10.2969/JMSJ/06941667], https://hal.archivesouvertes.fr/hal-01285311.
- [16] R. L. FRANK, F. MÉHATS, C. SPARBER. Averaging of nonlinear Schrödinger equations with strong magnetic confinement, in "Communications in Mathematical Sciences", 2017, vol. 15, n<sup>o</sup> 7, p. 1933-1945, https://arxiv. org/abs/1611.01574 - 12 pages [DOI: 10.4310/CMS.2017.v15.N7.A7], https://hal.archives-ouvertes.fr/hal-01397325.
- [17] M. LEMOU, A. M. LUZ, F. MÉHATS.Nonlinear stability criteria for the HMF Model, in "Archive for Rational Mechanics and Analysis", 2017, vol. 224, n<sup>o</sup> 2, p. 353-380, https://arxiv.org/abs/1509.08637 [DOI: 10.1007/s00205-017-1077-4], https://hal.archives-ouvertes.fr/hal-01207626.
- [18] M. LEMOU, F. MÉHATS, X. ZHAO. Uniformly accurate numerical schemes for the nonlinear dirac equation in the nonrelativistic limit regime, in "Communications in Mathematical Sciences", 2017, vol. 15, n<sup>o</sup> 4, p. 1107-1128 [DOI: 10.4310/CMS.2017.v15.N4.A9], https://hal.archives-ouvertes.fr/hal-01313976.
- [19] F. MÉHATS, O. PINAUD.*The quantum Liouville-BGK equation and the moment problem*, in "Journal of Differential Equations", 2017, vol. 263, n<sup>o</sup> 7, p. 3737-3787, https://arxiv.org/abs/1512.01504 [DOI: 10.1016/J.JDE.2017.05.004], https://hal.archives-ouvertes.fr/hal-01255137.

- [20] F. MÉHATS, N. RAYMOND.Strong confinement limit for the nonlinear Schrödinger equation constrained on a curve, in "Annales Henri Poincaré", 2017, vol. 18, n<sup>o</sup> 1, p. 281-306 [DOI: 10.1007/s00023-016-0511-8], https://hal.archives-ouvertes.fr/hal-01090045.
- [21] T. WANG, X. ZHAO, J. JIANG.Unconditional and optimal H<sup>2</sup>-error estimates of two linear and conservative finite difference schemes for the Klein-Gordon-Schrödinger equation in high dimensions, in "Advances in Computational Mathematics", 2017 [DOI: 10.1007/s10444-017-9557-5], https://hal.archives-ouvertes.fr/ hal-01576947.
- [22] X. ZHAO.A combination of multiscale time integrator and two-scale formulation for the nonlinear Schrödinger equation with wave operator, in "Journal of Computational and Applied Mathematics", 2017, vol. 326, p. 320 336 [DOI: 10.1016/J.CAM.2017.06.006], https://hal.archives-ouvertes.fr/hal-01576629.
- [23] X. ZHAO. Uniformly accurate multiscale time integrators for second order oscillatory differential equations with large initial data, in "BIT Numerical Mathematics", 2017, vol. 57, n<sup>o</sup> 3, p. 649 - 683 [DOI: 10.1007/s10543-017-0646-0], https://hal.archives-ouvertes.fr/hal-01591333.

#### Scientific Books (or Scientific Book chapters)

[24] M. CHAUSSADE-BEAUDOUIN, M. DAUGE, E. FAOU, Z. YOSIBASH.*High frequency oscillations of first eigenmodes in axisymmetric shells as the thickness tends to zero*, in "Operator Theory Advances and Application", Recent Trends in Operator Theory and Partial Differential Equations - The Roland Duduchava Anniversary Volume, Birkhäuser/Springer, 2017, vol. 258, p. 89-110, https://arxiv.org/abs/1603.01459 [DOI: 10.1007/978-3-319-47079-5\_5], https://hal.archives-ouvertes.fr/hal-01278861.

#### **Other Publications**

- [25] J. BERNIER. Optimality and resonances in a class of compact finite difference schemes of high order, October 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01612326.
- [26] C.-E. BRÉHIER, A. DEBUSSCHE. Kolmogorov Equations and Weak Order Analysis for SPDES with Nonlinear Diffusion Coefficient, 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01481966.
- [27] S. CERRAI, A. DEBUSSCHE. Large deviations for the dynamic  $\Phi_d^{2n}$  model, May 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01518465.
- [28] P. CHARTIER, M. LEMOU, F. MÉHATS, G. VILMART. A new class of uniformly accurate numerical schemes for highly oscillatory evolution equations, December 2017, working paper or preprint, https://hal.archivesouvertes.fr/hal-01666472.
- [29] A. CRESTETTO, N. CROUSEILLES, M. LEMOU.A particle micro-macro decomposition based numerical scheme for collisional kinetic equations in the diffusion scaling, January 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01439288.
- [30] N. CROUSEILLES, L. EINKEMMER, M. PRUGGER. *An exponential integrator for the drift-kinetic model*, 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01538450.
- [31] N. CROUSEILLES, S. JIN, M. LEMOU, L. LIU. Nonlinear Geometric Optics Based Multiscale Stochastic Galerkin Methods for Highly Oscillatory Transport Equations with Random Inputs \*, April 2017, working paper or preprint, https://hal.inria.fr/hal-01501825.
- [32] A. DE BOUARD, A. DEBUSSCHE, R. FUKUIZUMI.Long time behavior of Gross-Pitaevskii equation at positive temperature, 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01579123.
- [33] A. DEBUSSCHE, J. MARTIN. Solution to the stochastic Schrödinger equation on the full space, 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01579115.
- [34] A. DEBUSSCHE, M. J. NGUEPEDJA NANKEP.A Law of Large Numbers in the Supremum Norm for a Multiscale Stochastic Spatial Gene Network, November 2017, working paper or preprint, https://hal.archivesouvertes.fr/hal-01636039.
- [35] A. DEBUSSCHE, J. VOVELLE. Approximation diffusion in stochastically forced kinetic equations, July 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01567138.
- [36] M. FONTAINE, M. LEMOU, F. MÉHATS. Stable Ground States for the HMF Poisson Model, September 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01582008.

#### **References in notes**

- [37] E. HAIRER. Geometric integration of ordinary differential equations on manifolds, in "BIT", 2001, vol. 41, p. 996–1007.
- [38] E. HAIRER, C. LUBICH, G. WANNER. Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations, Second edition, Springer Series in Computational Mathematics 31, Springer, Berlin, 2006.
- [39] E. HAIRER, G. WANNER. Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems, Springer Series in Computational Mathematics 14, 2, Springer-Verlag, Berlin, 1996.
- [40] C. LUBICH.A variational splitting integrator for quantum molecular dynamics, in "Appl. Numer. Math.", 2004, vol. 48, p. 355–368.
- [41] C. LUBICH. On variational approximations in quantum molecular dynamics, in "Mathematics of Computation", 2009.
- [42] J. M. SANZ-SERNA, M. P. CALVO. Numerical Hamiltonian Problems, Chapman & Hall, London, 1994.

# **Project-Team KERDATA**

# Scalable Storage for Clouds and Beyond

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: Institut national des sciences appliquées de Rennes

Université Rennes 1

École normale supérieure de Rennes

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Distributed and High Performance Computing

### **Table of contents**

1.	Personnel	509	
2.	Overall Objectives		
	2.1.1. Our objective		
	2.1.1.1. Alignment with Inria's scientific strategy	510	
	2.1.1.2. Challenges and goals related to cloud data storage and processing	510	
	2.1.1.3. Challenges and goals related to data-intensive HPC applications	511	
	2.1.2. Our approach	511	
	2.1.2.1. Platforms and Methodology	511	
	2.1.2.2. Collaboration strategy	511	
3.	Research Program	511	
	3.1. Research axis 1: Convergence of Extreme-Scale Computing and Big Data Infrastructures	511	
	3.1.1. High-performance storage for concurrent Big Data applications	512	
	3.1.2. Big Data analytics on Exascale HPC machines	512	
	3.2. Research axis 2: Advanced data processing on Clouds	513	
	3.2.1. Stream-oriented, Big Data processing on clouds	513	
	3.2.2. Geographically distributed workflows on multi-site clouds	513	
	3.3. Research axis 3: I/O management, in situ visualization and analysis on HPC systems at extr	eme	
	scales	514	
4.	Highlights of the Year	514	
5.	New Software and Platforms	515	
	5.1. BlobSeer	515	
	5.2. Damaris	515	
	5.3. iHadoop	516	
	5.4. JetStream	516	
	5.5. OverFlow	516	
6.	New Results	516	
	6.1. Convergence of HPC and Big Data	516	
	6.1.1. Týr: Blob-based storage convergence of HPC and Big Data	516	
	6.1.2. Modeling elastic storage	517	
	6.1.3. Eley: Leveraging burst-buffers for efficient Big Data processing on HPC systems	517	
	6.2. Scalable data processing on clouds	518	
	6.2.1. Low-latency storage for stream processing	518	
	6.2.2. A Performance Evaluation of Apache Kafka in Support of Big Data Streaming App	lica-	
	tions	518	
	6.2.3. Hot metadata management for geographically distributed workflows	518	
	6.3. Scalable I/O, storage and in-situ processing in Exascale environments	519	
	6.3.1. Extreme-scale logging through application-defined storage	519	
	6.3.2. Leveraging Damaris for in-situ visualization in support of GeoScience and CFD Sim	ula-	
	tions	519	
	6.3.3. Accelerating MPI collective operations on the Theta cupercomputer	520	
	6.4. Energy-aware data storage and processing at large scale	520	
	6.4.1. Performance and energy-efficiency trade-offs in in-memory storage systems	520	
	6.4.2. Energy-aware straggler mitigation in Map-Reduce	521	
7.	Bilateral Contracts and Grants with Industry	521	
	7.1.1. Huawei: HIRP Low-Latency Storage for Stream Data (2016–2017)	521	
	7.1.2. Total: In situ Visualization with Damaris (2017-2018).	521	
8.	Partnerships and Cooperations	521	
	8.1. National Initiatives	521	
	8.1.1. ANR	521	

	8.1.1.1. OverFlow (2015–2019)	521	
	8.1.1.2. KerStream (2017–2021)	522	
	8.1.2. Other National Projects	522	
	8.1.2.1. ADT Damaris	522	
	8.1.2.2. Grid'5000	522	
	8.2. European Initiatives	523	
	8.2.1. FP7 & H2020 Projects	523	
	8.2.2. Collaborations with Major European Organizations	523	
	8.3. International Initiatives	523	
	8.3.1. Inria International Labs	523	
	8.3.2. Inria International Partners	524	
	8.3.2.1. Declared Inria International Partners	524	
	8.3.2.2. DataCloud@Work	524	
	8.3.2.3. Informal International Partners	525	
	8.3.3. Participation in Other International Programs	525	
	8.4. International Research Visitors	525	
	8.4.1. Visits of International Scientists	525	
	8.4.2. Visits to International Teams	525	
9.	Dissemination		
	9.1. Promoting Scientific Activities	525	
	9.1.1. Scientific Events Organisation	525	
	9.1.2. Scientific Events Selection	526	
	9.1.2.1. Chair of Conference Program Committees	526	
	9.1.2.2. Member of Conference Program Committees	526	
	9.1.2.3. Reviewer	526	
	9.1.3. Journal	526	
	9.1.3.1. Member of the Editorial Boards	526	
	9.1.3.2. Reviewer - Reviewing Activities	526	
	9.1.4. Keynote faiks and invited faiks	520	
	9.1.5. Leadership within the Scientific Community	527	
	9.1.0. Scientific Experiise	527	
	9.1.7. Research Administration	527	
	0.2.1 Teaching	527	
	0.2.2. Supervision	528	
	0.2.3 Juries	520	
	0.2.4 Miscellaneous	529	
	9.2.4 Responsibilities	529	
	9.2.7.1 Responsionates 9.2.4.2 Tutorials	529	
9.3 Popularization			
10.	Bibliography	529	
-0.	~~~~B-#k-1		

#### **Project-Team KERDATA**

Creation of the Team: 2009 July 01, updated into Project-Team: 2012 July 01

#### **Keywords:**

#### **Computer Science and Digital Science:**

- A1.1.4. High performance computing
- A1.1.5. Exascale
- A1.1.6. Cloud
- A1.1.9. Fault tolerant systems
- A1.3. Distributed Systems
- A1.6. Green Computing
- A3.1.2. Data management, quering and storage
- A3.1.3. Distributed data
- A3.1.8. Big data (production, storage, transfer)
- A6.2.7. High performance computing
- A6.3. Computation-data interaction
- A7.1. Algorithms
- A7.1.1. Distributed algorithms

#### **Other Research Topics and Application Domains:**

- B3.2. Climate and meteorology
- B3.3.1. Earth and subsoil
- B8.2. Connected city
- B9.4.5. Data science

# 1. Personnel

#### **Research Scientists**

Gabriel Antoniu [Team leader, Inria, Senior Researcher, HDR] Shadi Ibrahim [Inria, Researcher, until March 2017]

#### **Faculty Members**

Luc Bougé [École normale supérieure de Rennes, Professor, HDR] Alexandru Costan [INSA Rennes, Associate Professor]

#### **Technical Staff**

Hadi Salimi [Research Engineer, ADT Damaris, Inria]

#### **PhD Students**

Lokman Rahmani [University of Rennes 1, until February 2017, then a visitor from April 2017 to June 2017] Luis Eduardo Pineda Morales [Inria, until April 2017, currently R&D Engineer at ActiveEon] Tien-Dat Phan [University of Rennes 1, until November 2017, currently R&D Engineer at Dassault Systèmes] Orçun Yildiz [Inria, until November 2017] Ovidiu-Cristian Marcu [Inria] Mohammed-Yacine Taleb [Inria] Nathanaël Cheriere [École normale supérieure de Rennes] Paul Le Noac'h [INSA Rennes] Pierre Matri [Universidad Politécnica de Madrid (UPM), Espagne]

#### **Post-Doctoral Fellow**

Chi Zhou [Inria, until March 2017 (also known as Amelie Chi Zhou)]

#### Visiting Scientist

Jose Aguilar Canepa [Instituto Politécnico Nacional, Mexico, Mexique, from November 2017 to December 2017]

#### Administrative Assistant

Aurélie Patier [University of Rennes 1]

# 2. Overall Objectives

#### 2.1. Context: the need for scalable data management

We are witnessing a rapidly increasing number of application areas generating and processing very large volumes of data on a regular basis. Such applications are called *data-intensive*. Governmental and commercial statistics, climate modeling, cosmology, genetics, bio-informatics, high-energy physics are just a few examples in the scientific area. In addition, rapidly growing amounts of data from social networks and commercial applications are now routinely processed.

In all these examples, the overall application performance is highly dependent on the properties of the underlying data management service. It becomes crucial to store and manipulate massive data efficiently. However, these data are typically *shared* at a large scale and *concurrently accessed* at a high degree. With the emergence of recent infrastructures such as cloud computing platforms and post-Petascale high-performance computing (HPC) systems, achieving highly scalable data management under such conditions has become a major challenge.

#### 2.1.1. Our objective

The KerData project-team is namely focusing on designing innovative architectures and systems for *scal-able data storage and processing*. We target two types of infrastructures: *clouds* and *post-Petascale high-performance supercomputers*, according to the current needs and requirements of data-intensive applications.

We are especially concerned by the applications of major international and industrial players in cloud computing and extreme-scale high-performance computing (HPC), which shape the long-term agenda of the cloud computing [35], [32] and Exascale HPC [34] research communities. The Big Data area, which has recently captured a lot of attention, emphasized the challenges related to Volume, Velocity and Variety. This is yet another element of context that further highlights the primary importance of designing data management systems that are efficient at a very large scale.

#### 2.1.1.1. Alignment with Inria's scientific strategy

Data-intensive applications exhibit several common requirements with respect to the need for data storage and I/O processing. We focus on some core challenges related to data management, resulted from these requirements. Our choice is perfectly in line with Inria's strategic plan [39], which acknowledges as critical the challenges of *storing, exchanging, organizing, utilizing, handling and analyzing* the huge volumes of data generated by an increasing number of sources. This topic is also stated as a scientific priority of Inria's research centre of Rennes [38]: *Storage and utilization of distributed big data*.

2.1.1.2. Challenges and goals related to cloud data storage and processing

In the area of cloud data processing, a significant milestone is the emergence of the Map-Reduce [45] parallel programming paradigm. It is currently used on most cloud platforms, following the trend set up by Amazon [30]. At the core of Map-Reduce frameworks lies the storage system, a key component which must meet a series of specific requirements that are not fully met yet by existing solutions: the ability to provide efficient *fine-grain access* to the files, while sustaining a *high throughput* in spite of *heavy access concurrency*; the need to provide a high resilience to *failures*; the need to take *energy-efficiency* issues into account.

More recently, it becomes clear that data-intensive processing needs to go beyond the frontiers of single datacenters. In this perspective, extra challenges arise, related to the efficiency of metadata management. This efficiency has a major impact on the access to very large sets of small objects by Big Data processing workflows running on large-scale infrastructures.

#### 2.1.1.3. Challenges and goals related to data-intensive HPC applications

Key research fields such as climate modeling, solid Earth sciences or astrophysics rely on very largescale simulations running on post-Petascale supercomputers. Such applications exhibit requirements clearly identified by international panels of experts like IESP [37], EESI [33], ETP4HPC [34]. A jump of one order of magnitude in the size of numerical simulations is required to address some of the fundamental questions in several communities in this context. In particular, the lack of data-intensive infrastructures and methodologies to analyze the huge results of such simulations is a major limiting factor.

The challenge we have been addressing is to find new ways to store, visualize and analyze massive outputs of data during and after the simulations. Our main initial goal was to do it without impacting the overall performance, avoiding the *jitter* generated by I/O interference as much as possible. Recently, we started to focus specifically on *in situ processing* approaches and we explored approaches to *model and predict I/O phase occurrences* and to *reduce intra-application and cross-application I/O interference*.

#### 2.1.2. Our approach

KerData's global approach consists in studying, designing, implementing and evaluating distributed algorithms and software architectures for scalable data storage and I/O management for efficient, large-scale data processing. We target two main execution infrastructures: cloud platforms and post-Petascale HPC supercomputers.

#### 2.1.2.1. Platforms and Methodology

The highly experimental nature of our research validation methodology should be emphasized. To validate our proposed algorithms and architectures, we build software prototypes, then validate them at a large scale on real testbeds and experimental platforms.

We strongly rely on the Grid'5000 platform. Moreover, thanks to our projects and partnerships, we have access to reference software and physical infrastructures. In the cloud area, we use the Microsoft Azure and Amazon cloud platforms. In the post-Petascale HPC area, we are running our experiments on systems including some top-ranked supercomputers, such as Titan, Jaguar, Kraken or Blue Waters. This provides us with excellent opportunities to validate our results on advanced realistic platforms.

#### 2.1.2.2. Collaboration strategy

Our collaboration portfolio includes international teams that are active in the areas of data management for clouds and HPC systems, both in Academia and Industry.

Our academic collaborating partners include Argonne National Lab, University of Illinois at Urbana-Champaign, Universidad Politécnica de Madrid, Barcelona Supercomputing Center, University Politehnica of Bucharest. In industry, we are currently collaborating with Huawei and Total.

Moreover, the consortiums of our collaborative projects include application partners in the area of climate simulations (e.g., the Department of Earth and Atmospheric Sciences of the University of Michigan, within our collaboration inside JLESC [40]). This is an additional asset, which enables us to take into account application requirements in the early design phase of our solutions, and to validate those solutions with real applications... and real users!

# **3. Research Program**

### 3.1. Research axis 1: Convergence of Extreme-Scale Computing and Big Data Infrastructures

The tools and cultures of High Performance Computing and Big Data Analytics have evolved in divergent ways. This is to the detriment of both. However, big computations still generate and are needed to analyze Big Data. As scientific research increasingly depends on both high-speed computing and data analytics, the potential interoperability and scaling convergence of these two eco-systems is crucial to the future. Our objective for the next years is premised on the idea that we must begin to systematically map out and account for the ways in which the major issues associated with Big Data intersect with, impinge upon, and potentially change the plans that are now being laid for achieving Exascale computing.

**Collaboration.** This axis is addressed in close collaboration with María Pérez (UPM), Rob Ross (ANL), Toni Cortes (BSC), Bogdan Nicolae (formerly at IBM Research, now at Huawei Research).

Relevant groups with similar interests are the following ones.

- The group of Jack Dongarra, Innovative Computing Laboratory at University of Tennessee/Oak Ridge National Laboratory, working on joint tools Exascale Computing and Big Data.
- The group of Satoshi Matsuoka, Tokyo Institute of Technology, working on system software for Clouds and HPC.
- The group of Franck Cappello at Argonne National Laboratory/NCSA working on ondemand data analytics and storage for extreme-scale simulations and experiments.

#### 3.1.1. High-performance storage for concurrent Big Data applications

We argue that storage is a plausible pathway to convergence. In this context, we plan to focus on the needs of concurrent Big Data applications that require high-performance storage, as well as transaction support. Although blobs (binary large objects) are an increasingly popular storage model for such applications, state-of-the-art blob storage systems offer no transaction semantics. This demands users to coordinate data access carefully in order to avoid race conditions, inconsistent writes, overwrites and other problems that cause erratic behavior.

We argue there is a gap between existing storage solutions and application requirements, which limits the design of transaction-oriented applications. In this context, one idea on which we plan to focus our efforts is exploring how blob storage systems could provide built-in, multi-blob transactions, while retaining sequential consistency and high throughput under heavy access concurrency.

The early principles of this research direction have already raised interest from our partners at ANL (Rob Ross) and UPM (María Pérez) for potential collaborations. In this direction, the acceptance of our paper on the Týr transactional blob storage system as a Best Student Paper Award Finalist at the SC16 conference [10] is a very encouraging step.

#### 3.1.2. Big Data analytics on Exascale HPC machines

Big Data analytics is another interesting direction that we plan to explore, building on top of these converged storage architectures. Specifically, we will examine the ways in which Exascale infrastructures can be leveraged not only by HPC-centric, but also by scientific, cloud-centric applications. Many of the current state-of-the-art Big Data processing approaches, including Hadoop and Spark [41] are optimized to run on commodity machines. This impacts the mechanisms used to deal with failures and the limited network bandwidth.

A blind adoption of these systems on extreme-scale platforms would result in high overheads. It would therefore prevent users from fully benefiting from the high performance infrastructure. The objective that we set here is to explore design and implementation options for new data analytics systems that can exploit the features of extreme-scale HPC machines: multi-core nodes, multiple memory and storage technologies including a large memory, NVRAM, SSDs, etc.

#### 3.2. Research axis 2: Advanced data processing on Clouds

The recent evolutions in the area of Big Data processing have pointed out some limitations of the initial Map-Reduce model. It is well suited for batch data processing, but less suited for real-time processing of dynamic data streams. New types of data-intensive applications emerge, e.g., for enterprises who need to perform analysis on their stream data in ways that can give fast results (i.e., in real time) at scale (e.g., click-stream analysis and network-monitoring log analysis). Similarly, scientists require fast and accurate data processing techniques in order to analyze their experimental data correctly at scale (e.g., collectively analysis of large data sets distributed in multiple geographically distributed locations).

Our plan is to revisit current data management techniques to cope with the volatile requirements of dataintensive applications on large-scale dynamic clouds in a cost-efficient way.

**Collaboration.** This axis is addressed in close collaboration with María Pérez (UPM), Kate Keahey (ANL) and Toni Cortes (BSC).

Relevant groups with similar interests include the following ones.

- The AMPLab, UC Berkeley, USA, working on scheduling stream data applications in heterogeneous clouds.
- The group of *Ewa Deelman*, USC Information Sciences Institute, working on resource management for workflows in Clouds.
- The XTRA group, Nanyang Technological University, Singapore, working on resource provisioning for workflows in the cloud.

#### 3.2.1. Stream-oriented, Big Data processing on clouds

The state-of-the-art Hadoop Map-Reduce framework cannot deal with stream data applications, as it requires the data to be initially stored in a distributed file system in order to process them. To better cope with the abovementioned requirements, several systems have been introduced for stream data processing such as Flink [36], Spark [41], Storm [42], and Google MillWheel [44]. These systems keep computation in memory to decrease latency, and preserve scalability by using data-partitioning or dividing the streams into a set of deterministic batch computations.

However, they are designed to work in dedicated environments and they do not consider the performance variability (i.e., network, I/O, etc.) caused by resource contention in the cloud. This variability may in turn cause high and unpredictable latency when output streams are transmitted to further analysis. Moreover, they overlook the dynamic nature of data streams and the volatility in their computation requirements. Finally, they still address failures in a best-effort manner.

Our objective is to investigate new approaches for reliable, stream Big Data processing on clouds. We will explore new mechanisms that expose resource heterogeneity (observed variability in resource utilization at runtime) when scheduling stream data applications. We will also investigate how to adapt to node failures automatically, and to adapt the failure handling techniques to the characteristics of the running application and to the root cause of failures.

#### 3.2.2. Geographically distributed workflows on multi-site clouds

Many data processing jobs in data-intensive applications are modeled as workflows (i.e., as sets of tasks linked according to their data and computation dependencies) to facilitate the management and analysis of large volumes of data. With the fast growth of volumes of data to be handled at larger and larger scales, geographically distributed workflows are emerging as a natural data processing paradigm. This may bring several benefits: resilience to failures, distribution across partitions (e.g., moving computation close to data or vice versa), elastic scaling to support usage bursts, user proximity, etc.

In this context, sharing, disseminating and analyzing the data sets results in frequent large-scale data movements across widely distributed sites. Studies show that the inter-datacenter traffic is expected to triple in the following years. Our objective is to investigate approaches to data management enabling an efficient execution of such geographically distributed workflows running on multi-site clouds. While in the past years we have addressed some data management issues in this area, mainly in support to efficient task scheduling of scientific workflows running on multisite clouds, we will now focus on an increasingly common scenario where workflows generate and process a huge number of small files, which is particularly challenging. As such workloads generate a deluge of small and independent I/O operations, efficient data and metadata handling is critical. We will explore specific means to better hide latency for data and metadata access in such scenarios, as a way to improve global performance.

# 3.3. Research axis 3: I/O management, in situ visualization and analysis on HPC systems at extreme scales

Over the past few years, the increasing amounts of data produced by large-scale simulations have motivated a shift from traditional offline data analysis to in situ analysis and visualization. In situ processing started by coupling a parallel simulation with an analysis or visualization library, to avoid the cost of writing data on storage and reading it back. Going beyond this simple pairwise tight coupling, complex analysis workflows today are graphs with one or more data sources and several interconnected analysis components.

**Collaboration.** This axis is worked out in close collaboration with Rob Ross (ANL), Tom Peterka (ANL), Matthieu Dorier (ANL), Toni Cortes (BSC), Bruno Raffin (Inria). Some additional collaborations are in discussion with other members of JLESC, and with CEA and Total.

Relevant groups with similar interests include the following ones.

- The group of Manish Parashar at Rutgers University, USA (I/O management for HPC systems, in situ processing).
- The group of Scott Klasky at Oak Ridge National Lab, USA (I/O management for HPC systems, in situ processing).
- The CNRS IPSL laboratory (Sébastien Denvil, Pôle de modélisation du climat) in Paris, France (in situ data analytics).

#### 3.3.1. Toward a joint optimized architecture for in situ visualization and advanced processing

From Inria and ANL, four tools at least have emerged to address some challenges of coupling simulations with visualization packages or analysis workflows. Each of them focused on some particular aspect:

Damaris (Inria, [5], [4]) exploits dedicated cores to enable jitter-free I/O and in situ visualization;

Decaf (ANL, [31]) implements a coupling service for workflows;

FlowVR (Inria, [43]) connects workflow components for in situ processing;

Swift (ANL, [46]) focuses on implicitly parallel data flows and was optimized for Big Data processing.

Our plan is to explore how these tools could best leverage their respective strengths in a *joint optimized architecture for in situ visualization and advanced processing* in the HPC area. We published a preliminary study describing the lessons learned from using these tools in production environments with real applications [7]. Such a joint architecture will contribute to address the data volume and velocity challenges raised by dataintensive workflows, including complex data-intensive analytics phases. It may also impact, in a subsequent step, future data analysis pipelines for converged Big Data and HPC architectures.

# 4. Highlights of the Year

#### 4.1. Highlights of the Year

Euro-Par Steering Committee. Luc Bougé has been elected as the new Steering Commitee Chairman of the Euro-Par international conference on parallel and distributed processing. He is the successor of Prof. Christian Lengauer, University of Passau, Germany.

IEEE Cluster 2017 conference. Three years after the 2014 edition, the KerData team had again a leading role in the organization of the 2017 edition: Gabriel Antoniu served as Program Chair, Alexandru Costan served as Submissions Chair.

IEEE Big Data 2017 conference. Alexandru Costan served as Posters Chair.

# 5. New Software and Platforms

#### 5.1. BlobSeer

BlobSeer : A Storage System For The Exascale Era

KEYWORDS: Versioning - HPC - Cloud storage - Distributed metadata - MapReduce

SCIENTIFIC DESCRIPTION: BlobSeer is a large-scale distributed storage service that addresses advanced data management requirements resulting from ever-increasing data sizes. It is centered around the idea of leveraging versioning for concurrent manipulation of binary large objects in order to efficiently exploit data-level parallelism and sustain a high throughput despite massively parallel data access.

FUNCTIONAL DESCRIPTION: BlobSeer is a large-scale distributed storage service for advanced management of massive data. Validated on Nimbus, OpenNebula and Microsoft Azure cloud platforms.

- Participants: Bogdan Nicolae, Gabriel Antoniu and Luc Bougé
- Partners: Université de Rennes 1 ENS Cachan
- Contact: Gabriel Antoniu
- URL: http://blobseer.gforge.inria.fr/

#### 5.2. Damaris

KEYWORDS: Big data - Visualization - I/O - HPC - Exascale

SCIENTIFIC DESCRIPTION: Damaris is a middleware for multicore SMP nodes enabling them to efficiently handle data transfers for storage and visualization. The key idea is to dedicate one or a few cores of each SMP node to the application I/O. It is developed within the framework of a collaboration between KerData and the Joint Laboratory for Petascale Computing (JLPC). The current version enables efficient asynchronous I/O, hiding all I/O related overheads such as data compression and post-processing, as well as direct (in situ) interactive visualization of the generated data.

Damaris has been preliminarily evaluated at NCSA (Urbana-Champaign) with the CM1 tornado simulation code. CM1 is one of the target applications of the Blue Waters supercomputer in production at NCSA/UIUC (USA), in the framework of the Inria-UIUC-ANL Joint Lab (JLPC). Damaris now has external users, including (to our knowledge) visualization specialists from NCSA and researchers from the France/Brazil Associated research team on Parallel Computing (joint team between Inria/LIG Grenoble and the UFRGS in Brazil). Damaris has been successfully integrated into three large-scale simulations (CM1, OLAM, Nek5000). Works are in progress to evaluate it in the context of several other simulations including HACC (cosmology code) and GTC (fusion).

FUNCTIONAL DESCRIPTION: Damaris is a middleware for data management targeting large-scale HPC simulations: • «In-situ» data analysis by some dedicated cores of the simulation platform • Asynchronous and fast data transfer from HPC simulations to Damaris • Semantic-aware dataset processing through Damaris plug-ins

- Participants: Gabriel Antoniu, Lokman Rahmani, Luc Bougé, Matthieu Dorier and Orçun Yildiz
- Partner: ENS Rennes
- Contact: Matthieu Dorier
- URL: https://project.inria.fr/damaris/

### 5.3. iHadoop

FUNCTIONAL DESCRIPTION: iHadoop is a Hadoop simulator developed in Java on top of SimGrid to simulate the behavior of Hadoop and therefore accurately predict the performance of Hadoop in normal scenarios and under failures.

iHadoop is an internal software prototype, which was initially developed to validate our idea for exploring the behavior of Hadoop under failures. iHadoop has preliminarily evaluated within our group and it has shown very high accuracy when predicating the execution time of a Map-Reduce application. We intend to integrate iHadoop within the SimGrid distribution and make it available to the SimGrid community.

- Participants: Shadi Ibrahim and Tien Dat Phan
- Contact: Shadi Ibrahim

#### 5.4. JetStream

FUNCTIONAL DESCRIPTION: JetStream is a middleware solution for batch-based, high-performance streaming across cloud data centers. JetStream implements a set of context-aware strategies for optimizing batchbased streaming, being able to self-adapt to changing conditions. Additionally, the system provides multiroute streaming across cloud data centers for aggregating bandwidth by leveraging the network parallelism. It enables easy deployment across .Net frameworks and seamless binding with event processing engines such as StreamInsight.

JetStream is currently used at Microsoft Research ATLE Munich for the management of the Azure cloud infrastructure.

- Participants: Alexandru Costan, Gabriel Antoniu and Radu Marius Tudoran
- Contact: Alexandru Costan

#### 5.5. OverFlow

FUNCTIONAL DESCRIPTION: OverFlow is a uniform data management system for scientific workflows running across geographically distributed sites, aiming to reap economic benefits from this geo-diversity. The software is environment-aware, as it monitors and models the global cloud infrastructure, offering high and predictable data handling performance for transfer cost and time, within and across sites. OverFlow proposes a set of pluggable services, grouped in a data-scientist cloud kit. They provide the applications with the possibility to monitor the underlying infrastructure, to exploit smart data compression, deduplication and georeplication, to evaluate data management costs, to set a tradeoff between money and time, and optimize the transfer strategy accordingly.

Currently, OverFlow is used for data transfers by the Microsoft Research ATLE Munich team as well as for synthetic benchmarks at the Politehnica University of Bucharest.

- Participants: Alexandru Costan, Gabriel Antoniu and Radu Marius Tudoran
- Contact: Alexandru Costan

# 6. New Results

#### 6.1. Convergence of HPC and Big Data

#### 6.1.1. Týr: Blob-based storage convergence of HPC and Big Data

Participants: Pierre Matri, Alexandru Costan, Gabriel Antoniu.

The increasingly growing data sets processed on HPC platforms raise major challenges for the underlying storage layer. A promising alternative to POSIX-I/O-compliant file systems are simpler blobs (binary large objects), or object storage systems. They offer lower overhead, better performance and horizontal scalability at the cost of largely unused features such as file hierarchies or permissions. Similarly, blobs are increasingly considered for replacing distributed file systems for big data analytics or as a base for storage abstractions like key-value stores or time-series databases.

#### 516

This growing interest from both HPC and Big Data communities towards blob storage naturally fits with the current trend towards HPC and Big Data convergence. In this context, we seek to demonstrate that blob storage indeed constitutes a strong alternative to current storage infrastructures. Additionally, the data model of blob storage is close enough to that of distributed file systems so that this change is largely transparent for the applications running atop them.

In [22] we provide a preliminary evaluation of blob storage in HPC and Big Data contexts. We leverage a series of real-world HPC applications as well as an industry-standard HPC benchmark. We analyze for each of these applications the storage requests sent to the underlying storage system. We discover that over 98% of these storage calls can be directly mapped to the data model offered by blobs. Interestingly, we also note that the remaining calls are using file systems features for convinience rather than by necessity. These calls may consequently be performed as offline pre- or post-processing, or avoided altogether without altering the application.

#### 6.1.2. Modeling elastic storage

Participants: Nathanaël Cheriere, Gabriel Antoniu.

For efficient Big Data processing, efficient resource utilization becomes a major concern as large-scale computing infrastructures such as supercomputers or clouds keep growing in size. Naturally, energy and cost savings can be obtained by reducing idle resources. Malleability, which is the possibility for resource managers to *dynamically* increase or reduce the resources of jobs, appears as a promising means to progress towards this goal.

However, state-of-the-art parallel and distributed file systems have not been designed with malleability in mind. This is mainly due to the supposedly high cost of storage decommission, which is considered to involve expensive data transfers. Nevertheless, as network and storage technologies evolve, old assumptions on potential bottlenecks can be revisited.

In [18], we evaluate the viability of malleability as a design principle for a distributed file system. We specifically model the duration of the decommission operation, for which we obtain a theoretical lower bound. Then we consider HDFS as a use case and we show that our model can explain the measured decommission times.

The existing decommission mechanism of HDFS is good when the network is the bottleneck, but could be accelerated by up to a factor 3 when the storage is the limiting factor. With the highlights provided by our model, we suggest improvements to speed up decommission in HDFS and we discuss open perspectives for the design of efficient malleable distributed file systems.

#### 6.1.3. Eley: Leveraging burst-buffers for efficient Big Data processing on HPC systems

Participants: Orçun Yildiz, Chi Zhou, Shadi Ibrahim.

Burst Buffer is an effective solution for reducing the data transfer time and the I/O interference in HPC systems. Extending Burst Buffers (BBs) to handle Big Data applications is challenging because BBs must account for the large data inputs of Big Data applications and the performance guarantees of HPC applications – which are considered as first-class citizens in HPC systems. Existing BBs focus on only intermediate data of Big Data applications and incur a high performance degradation of both Big Data and HPC applications.In [26], we present *Eley*, a burst buffer solution that helps to accelerate the performance of Big Data applications while guaranteeing the performance of HPC applications. In order to improve the performance of Big Data applications to be stored close to computing nodes thus reducing the latency of reading data inputs. Moreover, Eley is equipped with a full delay operator to guarantee the performance of HPC applications – as they are running independently on a HPC system. The experimental results show the effectiveness of *Eley* in obtaining shorter execution time of Big Data applications (shorter map phase) while guaranteeing the performance of HPC applications.

#### 6.2. Scalable data processing on clouds

#### 6.2.1. Low-latency storage for stream processing

**Participants:** Ovidiu-Cristian Marcu, Alexandru Costan, Gabriel Antoniu, María Pérez, Radu Tudoran, Stefano Bortoli, Bogdan Nicolae.

We are now witnessing an unprecedented growth of data that needs to be processed at always increasing rates in order to extract valuable insights. Big Data applications are rapidly moving from a batch-oriented execution model to a streaming execution model in order to extract value from the data in real-time. Big Data streaming analytics tools have been developed to cope with the online dimension of data processing: they enable realtime handling of live data sources by means of stateful aggregations (window-based operators). In [21] we design a deduplication method specifically for window-based operators that rely on key-value stores to hold a shared state. Our key finding is that more fine-grained interactions between streaming engines and (key-value) stores (i.e., the data ingest, store, and process interfaces) need to be designed in order to better respond to scenarios that have to overcome memory scarcity.

Moreover, processing live data alone is often not enough: in many cases, such applications need to combine the live data with previously archived data to increase the quality of the extracted insights. Current streamingoriented runtimes and middlewares are not flexible enough to deal with this trend, as they address ingestion (collection and pre-processing of data streams) and persistent storage (archival of intermediate results) using separate services. This separation often leads to I/O redundancy (e.g., write data twice to disk or transfer data twice over the network) and interference (e.g., I/O bottlenecks when collecting data streams and writing archival data simultaneously). In [20] and [27] we argue for a unified ingestion and storage architecture for streaming data that addresses the aforementioned challenge and we identify a set of constraints and benefits for such a unified model, while highlighting the important architectural aspects required to implement it in real life.

Based on these findings, we are currently developing a low-latency stream storage framework that addresses such critical real-time needs for efficient stream processing, exposing high-performance interfaces for stream ingestion, storage, and processing.

#### 6.2.2. A Performance Evaluation of Apache Kafka in Support of Big Data Streaming Applications

#### Participants: Paul Le Noac'h, Alexandru Costan.

Stream computing is becoming a more and more popular paradigm as it enables the real-time promise of data analytics. Apache Kafka is currently the most popular framework used to ingest the data streams into the processing platforms. However, how to tune Kafka and how much resources to allocate for it remains a challenge for most users, who now rely mainly on empirical approaches to determine the best parameter settings for their deployments. Our goal in [28] is to make a through evaluation of several configurations and performance metrics of Kafka in order to allow users avoid bottlenecks, reach its full potential and avoid bottlenecks and eventually leverage some good practice for efficient stream processing.

#### 6.2.3. Hot metadata management for geographically distributed workflows

**Participants:** Luis Eduardo Pineda Morales, Alexandru Costan, Gabriel Antoniu, Ji Liu, Esther Pacitti, Patrick Valduriez, Marta Mattoso.

Large-scale scientific applications are often expressed as scientific workflows (SWfs) that help defining data processing jobs and dependencies between jobs' activities. Several SWfs have huge storage and computation requirements, and so they need to be processed in multiple (cloud-federated) datacenters. It has been shown that efficient metadata handling plays a key role in the performance of computing systems. However, most of this evidence concern only single-site, HPC systems to date. In addition, the efficient scheduling of tasks among different datacenters is critical to the SWf execution. In [19], we present a hybrid distributed model and architecture, using hot metadata (frequently accessed metadata) for efficient SWf scheduling in a multisite cloud. We couple our model with a scientific workflow management system (SWfMS) to validate

dynamically handle such hot metadata.

its applicability to real-life scientific workflows with different scheduling algorithms. We show that the combination of efficient management of hot metadata and scheduling algorithms improves the performance of SWfMS, reducing the execution time of highly parallel jobs up to 64.1 % and that of the whole scientific

#### 6.3. Scalable I/O, storage and in-situ processing in Exascale environments

#### 6.3.1. Extreme-scale logging through application-defined storage

Participants: Pierre Matri, Alexandru Costan, Gabriel Antoniu.

Applications generating data as logs and seeking to store it as such face hard challenges on HPC platforms. In distributed systems this storage model is key to ensuring fault-tolerance, developing transactional systems or publish-subscribe models. In scientific applications, distributed logs can play many roles such as in-situ visualization of large data streams, centralized collection of telemetry or monitoring events computational steering, data aggregation from array of physical sensors or live data indexing. Distributed shared logs are very difficult to implement on common HPC platforms due to the lack of efficient append operation in the current file-based storage infrastructures. While part of the POSIX standard, this operation has not been the main focus during the development of parallel file systems. While application-specific, custom-built solutions are possible, they require a significant development effort and often fail to meet the performance requirements of data-intensive applications running at large scale.

workflows up to 37.5 %, by avoiding unnecessary cold metadata operations. We also further discuss how to

In this work we go through the basic requirements of storing telemetry data streams for computational steering and visualization. For simple use cases where the telemetry data is only temporary, we prove that distributed logging can be performed at scale by leveraging state-of-the-art blob storage systems such as Týr or RADOS. This approach is supported by the growing availability of node-local storage on a new generation of supercomputers, giving application developers the freedom to deploy transient storage systems alongside the application directly on the compute nodes.

When long-term storage of the generated data is needed for offline visualization or analytics, we prove that distributed logs require a significantly lower number of output logs to achieve peak performance compared to Lustre or GPFS. We also prove that this low number of output files obviates the need for an explicit post-processing merge step in most cases for iterating the whole output log in generation order. We finally prove on up to 100,000 cores of the Theta supercomputer that our findings are applicable to run distributed logging at large scale, while improving write throughput by several orders of magnitude compared to Lustre of GPFS.

#### 6.3.2. Leveraging Damaris for in-situ visualization in support of GeoScience and CFD Simulations

Participants: Hadi Salimi, Matthieu Dorier, Luc Bougé.

Damaris is a middleware for in situ data analysis and visualization targeting extreme-scale, MPI-based simulations. The main goal of Damaris is to provide a simple method to instrument a simulation in order to benefit from in situ analysis and visualization. To this aim, the computing resources are partitioned such that a subset of cores in a SMP node or a subset of nodes of the underlying platform are dedicated to in situ processing. The data generated by the simulation are passed to these dedicated processes either through shared memory (in the case of dedicated cores) or through the MPI calls (in the case of dedicated nodes) and can be processed both in synchronous and asynchronous modes. Afterwards, the processed data can be analyzed or visualized. Damaris also supports a very simple API to instrument simulations developed in different domains. Moreover, using some XML configuration files for defining simulation data types (e.g. meshes) makes the instrumentation process easier with minimum code modifications. Active development is currently continuing within the KerData team, where it is at the center of several collaborations with industry (e.g Total) as well as with national and international academic partners.

In recent developments of Damaris, we have focused on two main targets that are: 1) Instrumenting new simulations codes from different scientific domains, i.e. geoscience and ocean modeling, 2) Implementing new storage backends, i.e. HDF5 for Damaris. In this regard, we report the results of some experiments we made to evaluate Damaris with respect to performance. These experiments were conducted on Grid'5000 test bed. In these experiments Damaris was employed to visualize the data generated by the Wave Propagation geoscience simulation and also the CROCO coastal and ocean simulation. During the experiments, the impact of Damaris was measured by comparing the simulations instrumented by Damaris (space partitioning approach) with a baseline where those simulations include data processing codes directly on their source code (time partitioning approach). The results of these simulations show that the incorporation of Damaris into a simulation decreases the total run time of the simulation due to its asynchronous data processing and visualization capabilities. In addition, using Damaris for data visualization has nearly no impact on the total run time of the mentioned simulation codes is much less compared to the case that the simulation code is instrumenting the simulation or storage APIs. Moreover, we also have studied the impact of new HDF5 storage backend, on storing simulation results in HDF5 format in both file-per-dedicated-core and collective I/O scenarios.

#### 6.3.3. Accelerating MPI collective operations on the Theta cupercomputer

Participants: Nathanaël Cheriere, Matthieu Dorier, Misbah Mubarak, Robert Ross, Gabriel Antoniu.

Recent network topologies in supercomputers have motivated new research on topology-aware collective communication algorithms for MPI. But such endeavor requires betting on the fact that topology-awareness is the primary factor to accelerate these collective operations. Besides, showing the benefit of a new, topology-aware algorithm requires not only access to a leadership-scale supercomputer with the desired topology, but also a large resource allocation on this supercomputer. Event-driven network simulations can alleviate such constraints and speed up the search for appropriate algorithms by providing early answers on their relative merit.

In our studies, we focus on the Scatter and AllGather operations in the context of the Theta supercomputer's dragonfly topology. We propose a set of topology-aware versions of these operations as well as optimizations of the old, non-topology-aware ones. We conduct an extensive simulation campaign using the CODES network simulator. Our results show that, contrary to our expectations, topology-awareness does not help improving significantly the speed of these operations. Rather, the high radix and low diameter of the dragonfly topology, along with already good routing protocols, enable simple algorithms based on non-blocking communications to perform better than state-of-the-art algorithms. A trivial implementation of Scatter using nonblocking point-to-point communications can be faster than state-of-the art algorithms by up to a factor of 6. Traditional AllGather algorithms can also be improved by the same principle and exhibit a 4x speedup in some situations. These results highlight the need to rethink the collective operations under the light of nonblocking communications.

#### 6.4. Energy-aware data storage and processing at large scale

#### 6.4.1. Performance and energy-efficiency trade-offs in in-memory storage systems

Participants: Mohammed-Yacine Taleb, Shadi Ibrahim, Gabriel Antoniu, Toni Cortes.

Most large popular web applications, like Facebook and Twitter, have been relying on large amounts of in-memory storage to cache data and offer a low response time. As the main memory capacity of clusters and clouds increases, it becomes possible to keep most of the data in the main memory. This motivates the introduction of in-memory storage systems. While prior work has focused on how to exploit the low-latency of in-memory access at scale, there is very little visibility into the energy-efficiency of in-memory storage systems. Even though it is known that main memory is a fundamental energy bottleneck in computing systems (i.e., DRAM consumes up to 40% of a server's power). During this project, by the means of experimental evaluation, we have studied the performance and energy-efficiency of RAMCloud - a well-known in-memory storage system. We reveal that although RAMCloud is scalable for read-only applications, it exhibits non-proportional power consumption. We also find that the current replication scheme implemented in RAMCloud

limits the performance and results in high energy consumption. Surprisingly, we show that replication can also play a negative role in crash-recovery.

#### 6.4.2. Energy-aware straggler mitigation in Map-Reduce

Participants: Tien-Dat Phan, Chi Zhou, Shadi Ibrahim, Guillaume Aupy, Gabriel Antoniu.

Energy consumption is an important concern for large-scale data-centers, which results in huge monetary cost for data-center operators. Due to the hardware heterogeneity and contentions between concurrent workloads, straggler mitigation is important to many Big Data applications running in large-scale data-centers and the speculative execution technique is widely-used to handle stragglers. Although a large number of studies have been proposed to improve the performance of Big Data applications using speculative execution, few of them have studied the energy efficiency of their solutions.

In [23], we propose two techniques to improve the energy efficiency of speculative executions while ensuring comparable performance. Specifically, we propose a hierarchical straggler detection mechanism which can greatly reduce the number of killed speculative copies and hence save the energy consumption. We also propose an energy-aware speculative copy allocation method which considers the trade-off between performance and energy when allocating speculative copies. We implement both techniques into Hadoop and evaluate them using representative Map-Reduce benchmarks. Results show that our solution can reduce the energy waste on killed speculative copies by up to 100% and improve the energy efficiency by 20% compared to state-of-the-art mechanisms.

# 7. Bilateral Contracts and Grants with Industry

#### 7.1. Bilateral Contracts with Industry

#### 7.1.1. Huawei: HIRP Low-Latency Storage for Stream Data (2016–2017)

Participants: Alexandru Costan, Ovidiu-Cristian Marcu, Gabriel Antoniu.

The goal of this project is to explore the plausible paths towards a dedicated storage solution for low-latency stream storage. Such a solution should provide on the one hand traditional storage functionality and on the other hand stream-like performance (i.e., low-latency I/O access to items and ranges of items).

We have investigated the main requirements and challenges, evaluated the different design choices (e.g., a standalone component vs. an extension of an existing Big Data solution like HDFS) and proposed a new converged architecture for smart storage.

#### 7.1.2. Total: In situ Visualization with Damaris (2017-2018).

Participants: Hadi Salimi, Matthieu Dorier, Gabriel Antoniu, Luc Bougé.

The goal of this expertise contract is to 1) disseminate the usage of Damaris for engineers at Total; 2) to realize a feasibility study for the usage of Damaris for in situ analysis of data for Total's HPC simulations.

# 8. Partnerships and Cooperations

#### 8.1. National Initiatives

#### 8.1.1. ANR

8.1.1.1. OverFlow (2015–2019)

Participants: Alexandru Costan, Paul Le Noac'h.

• Project Acronym: OverFlow.

- Project Title: Workflow Data Management as a Service for Multisite Applications.
- Coordinator: Alexandru Costan.
- Duration: Octobre 2015–October 2019.
- Other Partners: None (Young Researcher Project).
- External collaborators: Kate Keahey (University of Chicago and Argonne National Laboratory), Bogdan Nicolae (Huawei Research) and Christophe Blanchet (Institut Français de Bioinformatique).
- Abstract: This JCJC project led by Alexandru Costan investigates approaches to data management enabling an efficient execution of geographically distributed workflows running on multi-site clouds.
- Progress: In 2017, we have reviewed in depth the technical and architectural needs of data storage for the use cases that drive OverFlow, in order to consolidate a set of requirements for its future architecture. Based on these workflow traces, in a second step, we have investigated the suitable benchmarks that reasonably represent them. In this direction, we have first focused on ingestion and storage optimisations for such complex deployments, in particular the novel support for concurrent writes. The project was successfully reviewed at T0+18.

#### 8.1.1.2. KerStream (2017-2021)

Participant: Shadi Ibrahim.

- Project Acronym: KerStream.
- Project Title: Big Data Processing: Beyond Hadoop!
- Coordinator: Shadi Ibrahim .
- Duration: January 2017–January 2021.
- Other Partners: None (Young Researcher Project).
- Abstract: This JCJC project led by Shadi Ibrahim aims to address the limitations of Hadoop when running stream Big Data applications on large-scale clouds and to do a step beyond Hadoop by proposing a new approach, called KerStream, for scalable and resilient stream Big Data processing on clouds. The KerStream project can be seen as the first step towards developing the first French middleware that handles Stream Data processing at Scale.
- Note: Shadi Ibrahim left the KerData team in April 2017, so that this contract is no longer managed within the KerData team.

#### 8.1.2. Other National Projects

#### 8.1.2.1. ADT Damaris

Participants: Hadi Salimi, Alexandru Costan, Luc Bougé.

- Project Acronym: ADT Damaris
- Project Title: Technology development action for te Damaris environment.
- Coordinator: Alexandru Costan.
- Duration: 2016–2018.
- Abstract: This action aims to support the development of the Damaris software. Inria's *Technological Development Office* (D2T, *Direction du Dévelopment Technologique*) provided 2 years of funding support for a senior engineer.

Hadi Salimi is funded through this project to document, test and extend the Damaris software and make it a safely distributable product.

#### 8.1.2.2. Grid'5000

We are members of Grid'5000 community and run experiments on the Grid'5000 platform on a daily basis.

#### 8.2. European Initiatives

#### 8.2.1. FP7 & H2020 Projects

#### 8.2.1.1. BigStorage

Title: BigStorage: Storage-based Convergence between HPC and Cloud to handle Big Data. Programme: H2020.

Duration: January 2015–December 2018.

Coordinator: Universidad Politécnica de Madrid (UPM).

Partners:

- Barcelona Supercomputing Center Centro Nacional de Supercomputacion (Spain)
- CA Technologies Development Spain (Spain)
- CEA Commissariat à l'énergie atomique et aux énergies alternatives (France)
- Deutsches Klimarechenzentrum (Germany)
- Foundation for Research and Technology Hellas (Greece)
- Fujitsu Technology Solutions (Germany)
- Johannes Gutenberg Universitaet Mainz (Germany)
- Universidad Politecnica de Madrid (Spain)
- Seagate Systems UK (United Kingdom)
- Inria contact: Gabriel Antoniu and Adrien Lèbre.

URL: http://www.bigstorage-project.eu/.

Description: BigStorage is a European Training Network (ETN) whose main goal is to train future *data scientists*. It aims at enabling them and us to apply holistic and interdisciplinary approaches to take advantage of a data-overwhelmed world. This world requires *HPC* and *Cloud* infrastructures with a redefinition of *storage* architectures underpinning them — focusing on meeting highly ambitious performance and *energy* usage objectives. The KerData team is hosting 2 *Early Stage Researchers* in this framework and co-advises an extra PhD student.

#### 8.2.2. Collaborations with Major European Organizations

Gabriel Antoniu and Alexandru Costan are serving as Inria representatives in the working groups dedicated to *HPC-Big Data* convergence within the Big Data Value Association (BDVA) and the European Technology Platform in the area of High-Performance Computing (ETP4HPC). They are contributing to the respective Strategic Research Agendas of BDVA and ETP4HPC.

#### 8.3. International Initiatives

#### 8.3.1. Inria International Labs

8.3.1.1. JLESC: Joint Laboratory on Extreme-Scale Computing

The Joint Laboratory on Extreme-Scale Computing is jointly run by Inria, UIUC, ANL, BSC, JSC and RIKEN/AICS. It has been created in 2014 as a follow-up of the Inria-UIUC JLPC, the *Joint Laboratory for Petascale Computing*.

The KerData team is collaborating with teams from ANL and UIUC within this lab since 2009 on several topics in the areas of I/O, storage and in situ processing and cloud computing. This collaboration has been initially formalized as the *Data@Exascale* Associate Team with ANL and UIUC (2013–2015) followed by *Data@Exascale* 2 Associate Team with ANL (2016–2018). Our activities in this framework are described here: http://www.irisa.fr/kerdata/data-at-exascale/

Since 2015, Gabriel Antoniu serves as a topic leader for Inria for the *I/O*, *Storage and In Situ Processing* topic. Ongoing lab research directions and projects he is co-supervising in this area are described here: https://jlesc.github.io/projects/ in the *I/O*, *Storage and In-Situ Processing* section.

Since 2017, Gabriel Antoniu is serving as Vice-Executive Director of JLESC for Inria.

8.3.1.1.1. Associate Team involved in the International Lab: Data@Exascale 2

Project Acronym: Data@Exascale 2.

- Project Title: Convergent Data Storage and Processing Approaches for Exascale Computing and Big Data Analytics.
- International Partner: Argonne National Laboratory (United States) Mathematics and Computer Science Division (MCS) Rob Ross.

Start year: 2013.

URL: http://www.irisa.fr/kerdata/data-at-exascale/.

Description: In the past few years, countries including United States, the European Union, Japan and China have set up aggressive plans to get closer to what appears to be the next goal in terms of high-performance computing (HPC): Exaflop computing, a target which is now considered reachable by the next-generation supercomputers in 2020-2023. While these government-led initiatives have naturally focused on the big challenges of Exascale for the development of new hardware and software architectures, the quite recent emergence of the Big Data phenomenon introduces what could be called a tectonic shift that is impacting the entire research landscape for Exascale computing. As data generation capabilities in most science domains are now growing substantially faster than computational capabilities, causing these domains to become data-intensive, new challenges appeared in terms of volumes and velocity for data to be stored, processed and analyzed on the future Exascale machines.

To face the challenges generated by the exponential data growth (a general phenomenon in many fields), a certain progress has already been made in the recent years in the rapidly-developing, industry-led field of cloud-based Big Data analytics, where advanced tools emerged, relying on machine-learning techniques and predictive analytics.

Unfortunately, these advances cannot be immediately applied to Exascale computing: the tools and cultures of the two worlds, HPC (High-Performance Computing) and BDA (Big Data Analytics) have developed in a divergent fashion (in terms of major focus and technical approaches), to the detriment of both. The two worlds share however multiple similar challenges and unification now appears as essential in order to address the future challenges of major application domains that can benefit from both.

The scientific program we propose for the Data@Exascale 2 Associate Team is defined from this new, highly-strategic perspective and builds on the idea that the design of innovative approaches to data I/O, storage and processing allowing Big Data analytics techniques and the newest HPC architectures to leverage each other clearly appears as a key catalyst factor for the convergence process.

Activities in 2017 are described on the web site of the Associate Team.

#### 8.3.2. Inria International Partners

8.3.2.1. Declared Inria International Partners

8.3.2.2. DataCloud@Work

Title: DataCloud@Work.

International Partner:

 Polytechnic University of Bucharest (Romania), Computer Science Department, Nicolae Tapus and Valentin Cristea. Duration: 5 years.

Start year: 2013. The status of IIP was established right after the end of our former *DataCloud@work* Associate Team (2010–2012).

URL: https://www.irisa.fr/kerdata/doku.php?id=cloud\_at\_work:start.

- Description: Our research topics address the area of distributed data management for cloud services, focusing on autonomic storage. The goal is explore how to build an efficient, secure and reliable storage IaaS for data-intensive distributed applications running in cloud environments by enabling an autonomic behavior.
- 8.3.2.3. Informal International Partners
  - Instituto Politécnico Nacional, IPN, Ciudad de México: We continued our informal collaboration in the area of stream processing. A PhD student from IPN (José Aguilar Canepa) was hosted by the KerData team for a 1-month internship, during which he identified optimization problems that can be subject to joint work (see Internships section below).
  - National University of Singapore (NUS): We collaborate on resource management for workflows in the cloud and optimizing graph processing in geo-distributed data-centers.

#### 8.3.3. Participation in Other International Programs

- 8.3.3.1. International Initiatives
- 8.3.3.1.1. BDEC: Big Data and Extreme Computing

Since 2015, Gabriel Antoniu has been invited to participate to the yearly workshops of the international Big Data and Extreme-scale Computing (BDEC) working group, focused on the convergence of Extreme Computing (the latest incarnation of High-Performance Computing - HPC) and Big Data. BDEC is organized as an yearly series of invitation-based international workshops. In 2017 Gabriel Antoniu was solicited to colead the BDEC working group dedicated to exploring convergence-related challenges for hybrid architectures combining HPC systems, clouds and fog/edge computing infrastructures with Geoffrey Fox and Ewa Deelman. The contributions are reflected in the final report on convergence available on the BDEC web page.

#### 8.4. International Research Visitors

#### 8.4.1. Visits of International Scientists

José Aguilar Canepa (Instituto Politécnico Nacional, IPN, Mexico) visited the KerData team for one month (November 2017) in order to setup a common topic of research for the future proposal of an Associate Team Kerdata-IPN.

#### 8.4.1.1. Internships

Mukrram Rahman (M1, University of Rennes 1) has done a 3-month internship within the team, working with Ovidiu Marcu and Alexandru Costan on HDFS extensions for dedicated stream storage.

#### 8.4.2. Visits to International Teams

- 8.4.2.1. Research Stays Abroad
  - Pierre Matri has done a 3-month internship at Argonne National Lab, to work on extreme-scale logging through application-defined storage under the supervision of Phil Carns and Rob Ross. See Section *New Results* for details.

# 9. Dissemination

#### 9.1. Promoting Scientific Activities

#### 9.1.1. Scientific Events Organisation

9.1.1.1. General Chair, Scientific Chair

Luc Bougé: Chair of the Steering Committee of the Euro-Par Series of conferences since August 2017.

#### 9.1.2. Scientific Events Selection

9.1.2.1. Chair of Conference Program Committees

Gabriel Antoniu:

- Program Chair of the IEEE Cluster 2017 international conference.
- Vice-Chair of the Program Committee of the ACM/IEEE CCGrid 2017 international conference (Hybrid and Mobile Clouds Area), Madrid, May 2017.

Alexandru Costan:

- Program Co-Chair of the ScienceCloud 2017 international workshop held in conjunction with HPDC 2017, Washington, USA.
- Posters Chair of the IEEE Big Data 2017.
- Submissions Chair of the IEEE Cluster 2017.
- Track Chair of the IEEE ScalCom 2017 (Tools for Big Data track)

#### 9.1.2.2. Member of Conference Program Committees

Gabriel Antoniu: IEEE/ACM SC'17 (Papers), ACM HPDC 2017.

Luc Bougé: ISC HPC 2017.

Alexandru Costan: IEEE/ACM SC'17 (Posters), ACM/IEEE CCGrid 2017, IEEE/ACM UCC 2017, ARMS-CC 2017 workshop (held in conjunction with PODC 2016), IEEE Big Data 2017, MLDS 2017, EBDMA 2017, ISPDC 2017, CSCS 2017.

#### 9.1.2.3. Reviewer

Luc Bougé: ISC 2017, SC 2017, BigData 2017, IPDPS 2017, etc.

Alexandru Costan: HPDC 2017, SC 2017, IPDPS 2017

#### 9.1.3. Journal

#### 9.1.3.1. Member of the Editorial Boards

Gabriel Antoniu: Future Generation Computer Systems: Special Issue on Mobile, hybrid, and heterogeneous clouds for cyberinfrastructures (Guest Editor, 2017).

- Luc Bougé: Concurrency and Computation: Practice and Experience, Special Issues on the Euro-Par conference.
- Alexandru Costan: Soft Computing Journal, Special Issue on Autonomic Computing and Big Data Platforms

#### 9.1.3.2. Reviewer - Reviewing Activities

Gabriel Antoniu: Concurrency and Computation: Practice and Experience.

Luc Bougé: IEEE Transactions on Distributed Parallel Systems.

Alexandru Costan: IEEE Transactions on Parallel and Distributed Systems, Future Generation Computer Systems, Concurrency and Computation Practice and Experience, IEEE Communications, IEEE Transactions on Storage, Information Sciences, IEEE Transactions on Big Data.

Chi Zhou: IEEE Transactions on Big Data, IEEE Transactions on Cloud Computing and ACM TAAS.

#### 9.1.4. Keynote Talks and Invited Talks

Gabriel Antoniu:

- BigStorage and WALL ITN Joint Meeting: *Týr: Storage-based Convergence Between HPC and Big Data*, Mainz, January 2017.
- Huawei Seminar: *Convergence of HPC and Big Data*, Huawei European Research Center, Munich, October 2017.

Luc Bougé:

- Dagstuhl School on Challenges and Opportunities of User-Level File Systems for HPC: Are objects the right level of abstraction to enable the convergence between HPC and Big Data at storage level?, joint talk with María Pérez, Universidad Politécnica de Madrid (slides).
- ENS Rennes: *Map-Reduce: Very-Large Scale Programming for Big Data on Clouds*, November 2017.

Alexandru Costan:

- Huawei Seminar: Low-Latency Stream Storage, Huawei Research Munich, October 2017.
- BigStorage Summer School: Making Cities Smarter: A Storage-based View on Stream Processing Engines, Heraklion, July 2017.
- UPB Scientific Days: Science Driven, Scalable Data-Intensive Processing on Clouds, University Politehnica of Bucharest, June 2017.

#### 9.1.5. Leadership within the Scientific Community

Gabriel Antoniu:

- Scientific leader of the KerData project-team.
- Topic leader for Inria for the *Data storage*, *I/O and in situ processing* topic, supervising collaboration activities in this area within the JLESC, Joint Inria-Illinois-ANL-BSC-JSC-RIKEN/AICS Laboratory for Extreme-Scale Computing.
- Co-leader the BDEC working group dedicated to exploring convergence-related challenges for hybrid architectures combining HPC systems, clouds and fog/edge computing infrastructures with Geoffrey Fox and Ewa Deelman.
- Work package leader within the BigStorage H2020 ETN project for the *Data Science* work package.
- Luc Bougé: Vice-President of the *French Society for Informatics* (SIF), in charge of the Teaching department.
- Alexandru Costan: Leader of the *Smart Cities* Working Group within the **BigStorage** H2020 ETN project.

#### 9.1.6. Scientific Expertise

- Gabriel Antoniu served as a project evaluator for the project proposals submitted within the Activity "Post-doctoral Research Support" provided from the European Regional Development Fund to the The State Education Development Agency (SEDA) of the Republic of Latvia.
- Luc Bougé: Member of the jury for the *Agrégation de mathématiques* and the *CAPES of mathématiques*. These national committees select permanent mathematics teachers for secondary schools and high-schools, respectively.

#### 9.1.7. Research Administration

Gabriel Antoniu: Vice Executive Director of JLESC, Joint Inria-Illinois-ANL-BSC-JSC-RIKEN/AICS Laboratory for Extreme-Scale Computing for Inria.

#### 9.2. Teaching - Supervision - Juries

#### 9.2.1. Teaching

Gabriel Antoniu

Master (Engineering Degree, 5th year): Big Data, 24 hours (lectures), M2 level, ENSAI (École nationale supérieure de la statistique et de l'analyse de l'information), Bruz, France.

- Master : Cloud Computing, 15 hours (lectures and lab sessions), M2 level, ENSAI (École nationale supérieure de la statistique et de l'analyse de l'information), Bruz, France.
- Master: Scalable Distributed Systems, 12 hours (lectures), M1 level, SDS Module, EIT ICT Labs Master School, France.
- Master: Infrastructures for Big Data, 12 hours (lectures), M2 level, IBD Module, SIF Master Program, University of Rennes, France.
- Master: Cloud Computing and Big Data, 10 hours (lectures), M2 level, Cloud Module, MIAGE Master Program, University of Rennes, France.
- Master: Big Data Processing, 6 hours (lectures), M2 level, Health Big Data Master, University of Rennes, Faculty of Medecine, France.

#### Luc Bougé

- Bachelor: Introduction to programming concepts, 36 hours (lectures), L3 level, Informatics program, ENS Rennes, France.
- Master: Introduction to compilation, 24 hours (exercice and practical classes), M1 level, Informatics program, Univ. Rennes I, France.

#### Alexandru Costan

- Bachelor: Software Engineering and Java Programming, 28 hours (lab sessions), L3, INSA Rennes.
- Bachelor: Databases, 68 hours (lectures and lab sessions), L2, INSA Rennes, France.
- Bachelor: Practical case studies, 24 hours (project), L3, INSA Rennes.
- Master: Big Data Storage and Processing, 28h hours (lectures, lab sessions), M1, INSA Rennes.
- Master: Algorithms for Big Data, 28h hours (lectures, lab sessions), M2, INSA Rennes.
- Master: Big Data Project, 28h hours (project), M2, INSA Rennes.

#### 9.2.2. Supervision

- Luis Eduardo Pineda Morales: *Efficient Big Data Management for Geographically Distributed Work-flows*, INSA Rennes, defended on May 24, 2017. Co-advised by Alexandru Costan and Gabriel Antoniu. Manuscript: https://tel.archives-ouvertes.fr/tel-01645434.
- Tien Dat Phan: *Energy-efficient Straggler Mitigation for Big Data Applications on the Clouds*, ENS Rennes, defended on November 30, 2017. Co-advised by Shadi Ibrahim and Luc Bougé.
- Orçun Yildiz: *Efficient Big Data Processing on Large Scale Shared Platforms: Managing I/Os and Failures*, ENS Rennes, defended on December 8, 2017. Co-advised by Shadi Ibrahim and Gabriel Antoniu.

#### 9.2.2.1. PhD in progress

- Pierre Matri: *Predictive Models for Big Data*, thesis started in March 2015, co-advised by María Pérez (Universidad Politécnica de Madrid) and Gabriel Antoniu.
- Mohammed-Yacine Taleb: *Energy-impact of data consistency management in Clouds and Beyond*, thesis started in August 2015, co-advised by Gabriel Antoniu and Toni Cortés (Barcelona Supercomputing Center).
- Ovidiu-Cristian Marcu: *Efficient data transfer and streaming strategies for workflow-based Big Data processing*, thesis started in October 2015, co-advised by Alexandru Costan and Gabriel Antoniu.
- Nathanaël Cheriere: *Resource Management and Scheduling for Big Data Applications in Large-scale Systems*, thesis started in September 2016, co-advised by Gabriel Antoniu and Matthieu Dorier.
- Paul Le Noac'h: *Workflow Data Management as a Service for Multi-Site Applications*, thesis started in November 2016, co-advised by Alexandru Costan and Luc Bougé.

#### 9.2.3. Juries

Gabriel Antoniu: Referee for the PhD thesis of Jonathan Martí at the Barcelona Supercomputing Center (March 2017).

#### 9.2.4. Miscellaneous

9.2.4.1. Responsibilities

Gabriel Antoniu: in charge of the IBD module of the SIF Master of Rennes and of the Cloud module of the MIAGE Master of the University of Rennes 1.

Luc Bougé: Co-ordinator between ENS Rennes and the Inria Research Center and the IRISA laboratory.

Luc Bougé: In charge of the Bachelor level (L3) and of the student seminar series at the Informatics Departement of ENS Rennes.

Alexandru Costan: In charge of communication at the Computer Science Department of INSA Rennes. Alexandru Costan: In charge of the organization of the IRISA D1 Department Seminars.

#### 9.2.4.2. Tutorials

Gabriel Antoniu gave a tutorial on in situ processing at the BigStorage Summer School in Heraklion (July 2017).

#### 9.3. Popularization

Gabriel Antoniu:

ORAP Forum, Paris: ORAP is a French national forum intended for all stakeholders in the HPC area (industry and academia). Invited talk: *Convergence of HPC and Big Data: a storage-oriented perspective* in March 2017.

Available resources: Video and Slides.

Plenary Meeting of Inria's Department for International and European Relations: *Convergence of HPC and Big Data: Where Are We Going?*, Inria, Rocquencourt, October 2017.

Luc Bougé:

Doctoral Program, Rennes. Invited presentation to the PhD students about *Preparing your* applications after your PhD (November 2017).

Master Program, Rennes. Invited presentation to the M2 students about *Informatics as a scientific activity: Toward a responsible research* (November 2017).

## **10. Bibliography**

#### Major publications by the team in recent years

- [1] N. CHERIERE, M. DORIER. Design and Evaluation of Topology-aware Scatter and AllGather Algorithms for Dragonfly Networks, November 2016, Supercomputing 2016, Poster, https://hal.inria.fr/hal-01400271.
- [2] A. COSTAN, R. TUDORAN, G. ANTONIU, G. BRASCHE. TomusBlobs: Scalable Data-intensive Processing on Azure Clouds, in "CCPE - Concurrency and Computation: Practice and Experience", May 2013, https://hal. inria.fr/hal-00767034.
- [3] B. DA MOTA, R. TUDORAN, A. COSTAN, G. VAROQUAUX, G. BRASCHE, P. J. CONROD, H. LEMAITRE, T. PAUS, M. RIETSCHEL, V. FROUIN, J.-B. POLINE, G. ANTONIU, B. THIRION. *Machine Learning Patterns for Neuroimaging-Genetic Studies in the Cloud*, in "Frontiers in Neuroinformatics", April 2014, vol. 8, https://hal.inria.fr/hal-01057325.

- [4] M. DORIER, G. ANTONIU, F. CAPPELLO, M. SNIR, L. ORF. Damaris: How to Efficiently Leverage Multicore Parallelism to Achieve Scalable, Jitter-free I/O, in "CLUSTER - IEEE International Conference on Cluster Computing", Beijing, China, IEEE, September 2012, https://hal.inria.fr/hal-00715252.
- [5] M. DORIER, G. ANTONIU, F. CAPPELLO, M. SNIR, R. SISNEROS, O. YILDIZ, S. IBRAHIM, T. PETERKA, L. ORF.Damaris: Addressing Performance Variability in Data Management for Post-Petascale Simulations, in "ACM Transactions on Parallel Computing", 2016, https://hal.inria.fr/hal-01353890.
- [6] M. DORIER, G. ANTONIU, R. ROSS, D. KIMPE, S. IBRAHIM.CALCioM: Mitigating I/O Interference in HPC Systems through Cross-Application Coordination, in "IPDPS - International Parallel and Distributed Processing Symposium", Phoenix, United States, May 2014, https://hal.inria.fr/hal-00916091.
- [7] M. DORIER, M. DREHER, T. PETERKA, G. ANTONIU, B. RAFFIN, J. M. WOZNIAK. Lessons Learned from Building In Situ Coupling Frameworks, in "ISAV 2015 - First Workshop on In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization (held in conjunction with SC15)", Austin, United States, November 2015 [DOI: 10.1145/2828612.2828622], https://hal.inria.fr/hal-01224846.
- [8] M. DORIER, S. IBRAHIM, G. ANTONIU, R. ROSS. Omnisc'10: A Grammar-Based Approach to Spatial and Temporal I/O Patterns Prediction, in "SC14 - International Conference for High Performance Computing, Networking, Storage and Analysis", New Orleans, United States, IEEE, ACM, November 2014, https://hal. inria.fr/hal-01025670.
- [9] M. DORIER, S. IBRAHIM, G. ANTONIU, R. ROSS. Using Formal Grammars to Predict I/O Behaviors in HPC: the Omnisc'IO Approach, in "TPDS - IEEE Transactions on Parallel and Distributed Systems", October 2015 [DOI: 10.1109/TPDS.2015.2485980], https://hal.inria.fr/hal-01238103.
- [10] P. MATRI, A. COSTAN, G. ANTONIU, J. MONTES, M. S. PÉREZ. Týr: Blob Storage Meets Built-In Transactions, in "IEEE ACM SC16 - The International Conference for High Performance Computing, Networking, Storage and Analysis 2016", Salt Lake City, United States, November 2016, https://hal.inria. fr/hal-01347652.
- [11] B. NICOLAE, G. ANTONIU, L. BOUGÉ, D. MOISE, A. CARPEN-AMARIE. BlobSeer: Next-Generation Data Management for Large-Scale Infrastructures, in "JPDC - Journal of Parallel and Distributed Computing", February 2011, vol. 71, n<sup>o</sup> 2, p. 169–184, http://hal.inria.fr/inria-00511414/en/.
- [12] B. NICOLAE, J. BRESNAHAN, K. KEAHEY, G. ANTONIU. Going Back and Forth: Efficient Multi-Deployment and Multi-Snapshotting on Clouds, in "HPDC 2011 - The 20th International ACM Symposium on High-Performance Parallel and Distributed Computing", San José, CA, United States, June 2011, http://hal.inria.fr/ inria-00570682/en.
- [13] R. TUDORAN, A. COSTAN, G. ANTONIU. OverFlow: Multi-Site Aware Big Data Management for Scientific Workflows on Clouds, in "IEEE Transactions on Cloud Computing", June 2015 [DOI: 10.1109/TCC.2015.2440254], https://hal.inria.fr/hal-01239128.

#### **Publications of the year**

#### **Doctoral Dissertations and Habilitation Theses**

[14] T.-D. PHAN. Energy-efficient Straggler Mitigation for Big Data Applications on the Clouds, ENS Rennes, November 2017, https://tel.archives-ouvertes.fr/tel-01669469.

- [15] L. E. PINEDA MORALES. Efficient support for data-intensive scientific workflows on geo-distributed clouds, INSA de Rennes, May 2017, https://tel.archives-ouvertes.fr/tel-01645434.
- [16] O. YILDIZ. Efficient Big Data Processing on Large-Scale Shared Platforms: Managing I/Os and Failures, ENS Rennes, December 2017, https://tel.archives-ouvertes.fr/tel-01671413.

#### **Articles in International Peer-Reviewed Journal**

[17] P. MATRI, M. S. PÉREZ, A. COSTAN, L. BOUGÉ, G. ANTONIU. Keeping up with storage: Decentralized, write-enabled dynamic geo-replication, in "Future Generation Computer Systems", June 2017, p. 1-19 [DOI: 10.1016/J.FUTURE.2017.06.009], https://hal.inria.fr/hal-01617658.

#### **International Conferences with Proceedings**

- [18] N. CHERIERE, G. ANTONIU. *How Fast Can One Scale Down a Distributed File System?*, in "BigData 2017 IEEE International Conference on Big Data", Boston, United States, December 2017, https://hal.archives-ouvertes.fr/hal-01644928.
- [19] J. LIU, L. PINEDA-MORALES, E. PACITTI, A. COSTAN, P. VALDURIEZ, G. ANTONIU, M. MAT-TOSO. Efficient Scheduling of Scientific Workflows using Hot Metadata in a Multisite Cloud, in "BDA: Conférence sur la Gestion de Données — Principes, Technologies et Applications", Nancy, France, November 2017, 13, https://hal-lirmm.ccsd.cnrs.fr/lirmm-01620231.
- [20] O.-C. MARCU, A. COSTAN, G. ANTONIU, M. S. PÉREZ-HERNÁNDEZ, R. TUDORAN, S. BORTOLI, B. NICOLAE. *Towards a Unified Storage and Ingestion Architecture for Stream Processing*, in "Second Workshop on Real-time & Stream Analytics in Big Data Colocates with the 2017 IEEE International Conference on Big Data", Boston, United States, IEEE, December 2017, p. 1-6, https://hal.inria.fr/hal-01649207.
- [21] O.-C. MARCU, R. TUDORAN, B. NICOLAE, A. COSTAN, G. ANTONIU, M. S. PÉREZ-HERNÁNDEZ. Exploring Shared State in Key-Value Store for Window-Based Multi-Pattern Streaming Analytics, in "Workshop on the Integration of Extreme Scale Computing and Big Data Management and Analytics in conjunction with IEEE/ACM CCGrid 2017", Madrid, Spain, May 2017, https://hal.inria.fr/hal-01530744.
- [22] P. MATRI, Y. ALFOROV, A. BRANDON, M. KUHN, P. CARNS, T. LUDWIG. Could Blobs Fuel Storage-Based Convergence Between HPC and Big Data?, in "CLUSTER 2017 - IEEE International Conference on Cluster Computing", Honolulu, United States, September 2017, p. 81 - 86 [DOI: 10.1109/CLUSTER.2017.63], https://hal.inria.fr/hal-01617655.
- [23] T.-D. PHAN, S. IBRAHIM, A. C. ZHOU, G. AUPY, G. ANTONIU. Energy-Driven Straggler Mitigation in MapReduce, in "Euro-Par'17 - 23rd International European Conference on Parallel and Distributed Computing", Santiago de Compostela, Spain, August 2017, https://hal.inria.fr/hal-01560044.
- [24] Y. TALEB, S. IBRAHIM, G. ANTONIU, T. CORTES. An Empirical Evaluation of How The Network Impacts The Performance and Energy Efficiency in RAMCloud, in "Workshop on the Integration of Extreme Scale Computing and Big Data Management and Analytics in conjunction with IEEE/ACM CCGrid 2017", Madrid, Spain, May 2017, https://hal.inria.fr/hal-01376923.

- [25] Y. TALEB, S. IBRAHIM, G. ANTONIU, T. CORTES. Characterizing Performance and Energy-Efficiency of The RAMCloud Storage System, in "The 37th IEEE International Conference on Distributed Computing Systems (ICDCS 2017)", Atlanta, United States, June 2017, https://hal.inria.fr/hal-01496959.
- [26] O. YILDIZ, A. C. ZHOU, S. IBRAHIM. Eley: On the Effectiveness of Burst Buffers for Big Data Processing in HPC systems, in "Cluster'17-2017 IEEE International Conference on Cluster Computing", Hawaii, United States, September 2017, https://hal.inria.fr/hal-01570737.

#### **Research Reports**

[27] O.-C. MARCU, A. COSTAN, G. ANTONIU, M. S. PÉREZ-HERNÁNDEZ.Kera: A Unified Storage and Ingestion Architecture for Efficient Stream Processing, Inria Rennes - Bretagne Atlantique, June 2017, n<sup>o</sup> RR-9074, https://hal.inria.fr/hal-01532070.

#### **Other Publications**

- [28] P. LE NOAC'H, A. COSTAN, L. BOUGÉ. A Performance Evaluation of Apache Kafka in Support of Big Data Streaming Applications, December 2017, IEEE Big Data 2017, Poster, https://hal.archives-ouvertes.fr/hal-01647229.
- [29] Y. TALEB, R. STUTSMAN, G. ANTONIU, T. CORTES. *Tailwind: fast and atomic RDMA-based replication*, January 2018, working paper or preprint, https://hal.inria.fr/hal-01676502.

#### **References in notes**

- [30] Amazon Elastic Map-Reduce (EMR), 2017, https://aws.amazon.com/emr/.
- [31] The Decaf Project, 2017, https://bitbucket.org/tpeterka1/decaf.
- [32] Digital Single Market, 2015, https://ec.europa.eu/digital-single-market/en/digital-single-market.
- [33] European Exascale Software Initiative, 2013, http://www.eesi-project.eu.
- [34] The European Technology Platform for High-Performance Computing, 2012, http://www.etp4hpc.eu.
- [35] European Cloud Strategy, 2012, https://ec.europa.eu/digital-single-market/en/european-cloud-computingstrategy.
- [36] Apache Flink, 2016, http://flink.apache.org.
- [37] International Exascale Software Program, 2011, http://www.exascale.org/iesp/Main\_Page.
- [38] Scientific challenges of the Inria Rennes-Bretagne Atlantique research centre, 2016, https://www.inria.fr/en/ centre/rennes/research.
- [39] Inria's strategic plan "Towards Inria 2020", 2016, https://www.inria.fr/en/institute/strategy/strategic-plan.
- [40] Joint Laboratory for Extreme Scale Computing (JLESC), 2017, https://jlesc.github.io.

- [41] Apache Spark, 2017, http://spark.apache.org.
- [42] Storm, 2014, http://storm-project.net/.
- [43] The FlowVR Project, 2014, http://flowvr.sourceforge.net/.
- [44] T. AKIDAU, A. BALIKOV, K. BEKIROĞLU, S. CHERNYAK, J. HABERMAN, R. LAX, S. MCVEETY, D. MILLS, P. NORDSTROM, S. WHITTLE.*MillWheel: fault-tolerant stream processing at internet scale*, in "Proceedings of the VLDB Endowment", 2013, vol. 6, n<sup>O</sup> 11, p. 1033–1044.
- [45] J. DEAN, S. GHEMAWAT. *MapReduce: simplified data processing on large clusters*, in "Communications of the ACM", 2008, vol. 51, n<sup>0</sup> 1, p. 107–113.
- [46] S. WILDE, M. HATEGAN, J. M. WOZNIAK, B. CLIFFORD, D. KATZ, I. T. FOSTER.Swift: A language for distributed parallel scripting, in "Parallel Computing", 2011, vol. 37, n<sup>o</sup> 9, p. 633–652, http://dx.doi.org/10. 1016/j.parco.2011.05.005.

# **Project-Team LACODAM**

# Large scale Collaborative Data Mining

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: Institut national des sciences appliquées de Rennes

Institut national supérieur des sciences agronomiques, agroalimentaires, horticoles et du paysage

Université Rennes 1

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Data and Knowledge Representation and Processing

# **Table of contents**

1.	Personnel				
2.	Overall Objectives				
3.	Research Program	539			
	3.1. Introduction	539			
	3.2. Pattern mining algorithms	540			
	3.3. User/system interaction	541			
	3.4. Decision support	542			
	3.5. Long-term goals	542			
4.	Application Domains	. 543			
	4.1. Introduction	543			
	4.2. Industry	543			
	4.3. Health	544			
	4.4. Agriculture and environment	545			
5.	Highlights of the Year	546			
6.	New Software and Platforms	. 546			
	6.1. EcoMata	546			
	6.2. PATURMATA	546			
	6.3. Promise	546			
	6.4. GWASDM	547			
	6.5. DCM	547			
	6.6. NTGSP	547			
	6.7. Relevant Interval Rules Miner	548			
	6.8. OCL	548			
7.	New Results	548			
	7.1. Introduction	548			
	7.1.1. Pattern Mining	548			
	7.1.2. Decision Support	549			
	7.2. Discriminant chronicles mining: Application to care pathways analytics	549			
	7.3. Purchase Signatures of Retail Customers	549			
	7.4. Topic Signatures in Political Campaign Speeches	549			
	7.5. Expert Opinion Extraction from a Biomedical Database	550			
	7.6. Mining Relevant Interval Rules	550			
	7.7. Time Series Rule Matching: Application to Energy Consumption	550			
	7.8. Negative Temporal Sequence Mining	550			
	7.9. TopPI: An efficient algorithm for item-centric mining	551			
	7.10. Declarative Sequential Pattern Mining of Care Pathways	551			
	7.11. Efficiency Analysis of ASP Encodings for Sequential Pattern Mining Tasks	551			
	7.12. Mining Rare Sequential Patterns with ASP	551			
	7.13. From Medico-administrative Databases Analysis to Care Trajectories Analytics: An Exan	aple			
	with the French SNDS	552			
	7.14. A Data warehouse to Explore Multidimensional Simulated Data from a Spatially Distribution				
	Agro-nydrological Model to Improve Catchment Nitrogen Management	552			
	7.15. Extended Automata for Temporal Planning of Interacting Agents	552			
0	(.10. Anomaly Detection in Streams with Extreme Value Theory <b>Bilatoral Contracts and Croats with Industry</b>	552			
ð.	9.1 Dilateral Contracts with Industry	550			
	8.1. TTD A MI. Interactive Trace Mining	552			
	8.1.2 Hyptser: Hybrid Drediction of Time Series	552			
	8.1.2. Inputson inputson of the Strategic Action EDE/Inria	552			
	6.1.5. I articular Contract of the Strategic Action EDI/Initia	555			

	8.2. B	lateral Grants with Industry	554
9.	Partnerships and Cooperations		554
	9.1. R	egional Initiatives	554
	9.2. N	ational Initiatives	554
	9.2.1.	ANR	554
	9.2.2.	National Platforms	555
	9.3. In	ternational Research Visitors	555
10.	Dissem	ination	555
10.1. Promoting Scientific Activities			555
	10.1.1	. Scientific Events Organisation	555
	10.1.2	. Scientific Events Selection	555
	10	1.2.1. Member of the Conference Program Committees	555
	10	1.2.2. Reviewer	556
	10.1.3	. Journal	556
	10	1.3.1. Member of the Editorial Boards	556
	10	1.3.2. Reviewer - Reviewing Activities	556
10.1.4. Invited Talks		. Invited Talks	556
	10.1.5	. Leadership within the Scientific Community	556
	10.1.6	. Scientific Expertise	557
	10.1.7	. Research Administration	557
	10.2. Te	eaching - Supervision - Juries	557
	10.2.1	. Teaching	557
	10.2.2	. Supervision	557
	10.2.3	. Juries	557
	10.3. Po	opularization	558
11.	Bibliog	raphy	558

### **Project-Team LACODAM**

*Creation of the Team: 2016 January 01, updated into Project-Team: 2017 November 01* **Keywords:** 

#### **Computer Science and Digital Science:**

- A2.1. Programming Languages
- A2.1.5. Constraint programming
- A2.1.11. Proof languages
- A3. Data and knowledge
- A3.1. Data
- A3.1.1. Modeling, representation
- A3.2. Knowledge
- A3.2.1. Knowledge bases
- A3.2.2. Knowledge extraction, cleaning
- A3.2.3. Inference
- A3.2.4. Semantic Web
- A3.3. Data and knowledge analysis
- A3.3.1. On-line analytical processing
- A3.3.2. Data mining
- A3.3.3. Big data analysis
- A3.4. Machine learning and statistics
- A3.4.1. Supervised learning
- A3.4.2. Unsupervised learning
- A3.4.6. Neural networks
- A3.4.8. Deep learning
- A4. Security and privacy
- A4.9.1. Intrusion detection
- A7.1. Algorithms
- A7.2. Logic in Computer Science
- A7.2.1. Decision procedures
- A7.2.2. Automated Theorem Proving
- A9. Artificial intelligence
- A9.1. Knowledge
- A9.2. Machine learning
- A9.3. Signal analysis
- A9.6. Decision support
- A9.7. AI algorithmics

#### **Other Research Topics and Application Domains:**

- B1.1.6. Genomics
- B2. Health
- B2.3. Epidemiology
- B2.4. Therapies

B2.4.2. - Drug resistance B3. - Environment and planet B3.4. - Risks B3.4.3. - Pollution B3.5. - Agronomy B3.6. - Ecology B3.6.1. - Biodiversity B4. - Energy B6. - IT and telecom B6.2. - Network technologies **B8.** - Smart Cities and Territories B8.1. - Smart building/home B8.1.1. - Energy for smart buildings B8.1.2. - Sensor networks for smart buildings B8.2. - Connected city B8.3. - Urbanism and urban planning B9. - Society and Knowledge B9.4.5. - Data science B9.9.1. - Environmental risks

# **1. Personnel**

#### **Research Scientists**

Luis Galárraga Del Prado [Inria, Researcher, from Oct. 2017] René Quiniou [Inria, Researcher] Torsten Schaub [University of Postdam, Inria International Chair, Researcher, HDR]

#### **Faculty Members**

Alexandre Termier [Team leader, Univ. Rennes I, Professor, Partial secondment Inria from September 2016, HDR]

Marie-Odile Cordier [Univ. Rennes I, Faculty Member, Emeritus Professor, HDR] Elisa Fromont [Univ. Rennes I, Faculty Member, Professor, from Sep 2017, HDR] Véronique Masson [Univ. Rennes I, Associate Professor, Partial secondment Inria from September 2017] Laurence Rozé [INSA Rennes, Associate Professor] Thomas Guyet [AGROCAMPUS OUEST, Associate Professor] Christine Largouët [AGROCAMPUS OUEST, Associate Professor]

#### **External Collaborators**

Philippe Besnard [CNRS, Researcher, HDR] Anne-Isabelle Graux [INRA, Researcher]

#### **Technical Staff**

Louis Bonneau de Beaufort [AGROCAMPUS OUEST, Engineer]

#### PhD Students

Yann Dauxais [Univ. Rennes I] Kevin Fauvel [Inria, from Oct. 2017] Clément Gautrais [Univ. de Rennes I] Maël Guillemé [Energiency, granted by CIFRE] Colin Leverger [Orange Labs, from Oct. 2017] Alban Siffer [Amossys, from May 2016]

#### **Post-Doctoral Fellow**

Ahmed Samet [Univ. Rennes I, until May 2017]

#### Administrative Assistant

Marie-Noëlle Georgeault [Inria]

# 2. Overall Objectives

#### 2.1. Overall Objectives

Data collection is ubiquitous nowadays and it is providing our society with tremendous volumes of knowledge about human, environmental and, industrial activity. This ever-increasing quantity of data holds the keys to new discoveries, both in industrial and scientific domains. However, those keys will only be accessible to those who can make sense out of such data. Making sense out of data is a hard problem. It requires a good understanding of the data at hand, proficiency with the available analysis tools and methods, and good deductive skills. All these skills have been grouped under the umbrella term "Data Science" and universities have put a lot of effort in producing professionals in this field. "Data Scientist" is currently the most sought-after job in the USA, as the demand far exceeds the number of competent professionals. Despite its boom, data science is still mostly a "manual" process: current data analysis tools still require a significant amount of human effort and know-how. This makes data analysis a lengthy and error-prone process. This is true even for data science experts, and current approaches are mostly out of reach of non-specialists.

We claim that nowadays, Data Science is in its "Iron Age": Good tools are available, however skilled craftsmen are required to use them in order to transform raw material (the data) into finished products (knowledge, decisions). We foresee that in a decade from now, we should be in an "Industrial Age" of Data Science, where more elaborate tools will alleviate a lot of the human work required in Data Science. Basic Data Science tasks will no longer require a skilled data scientist; instead software tools will enable small companies or even individuals to get valuable knowledge from their data. Skilled data scientists will thus be fully available to work on the hard tasks that matter. This will entail a drastic improvement in productivity thanks to a new generation of tools that will do the tedious work for data analysts and scientists.

The objective of the LACODAM team is to facilitate the process of making sense out of large amounts of data. This can serve the purpose of deriving knowledge and insights for better decision-making. Since data science in its current state involves lots of human intervention, we envision a novel generation of data analysis and decision support tools that require significantly less tedious human work. Such solutions will rely only on a few interactions between the user and the system with high added value. We foresee solutions that bridge data mining techniques with artificial intelligence (AI) approaches, in order to integrate existing automated reasoning techniques in knowledge discovery workflows. Such solutions can be seen as "second order" AI tasks: they exploit AI techniques (for example, planning) in order to pilot more classical AI tasks such as data mining and decision support.

# 3. Research Program

#### **3.1. Introduction**

The three research axes of the LACODAM project-team are the following. First, we briefly introduce these axes, as well as their interplay:

• The first research axis is dedicated to the design of *novel pattern mining methods*. Pattern mining is one of the most important approaches to discover novel knowledge in data, and one of our strongest areas of expertise. The work on this axis will serve as foundations for work on the other two axes. Thus, this axis will have the strongest impact on our goals overall.

- The second axis tackles another aspect of knowledge discovery in data: the *interaction between the user and the system* in order to co-discover novel knowledge. Our team has plenty of experience collaborating with domain experts, and is therefore aware of the need to improve such interaction.
- The third axis concerns *decision support*. With the help of methods from the two previous axes, our goal here is to design systems that can either assist humans with making decisions, or make relevant decisions in situations where extremely fast reaction is required.

The following figure sums up the detailed work presented in the next few pages: we show the three research axes of the team (X-axis) on the left and our main applications areas (Y-axis) below. In the middle there are colored squares that represent the precise research topics of the team aligned with their axis and main application area. These research topics will be described in this section. Lines represent projects that can link several topics, and that are also connected to their main application area.



Lacodam research focus seen through its short term thematic applications

Figure 1. Lacodam research topics organized by axis and application

#### 3.2. Pattern mining algorithms

Twenty years of research in pattern mining have resulted in efficient approaches to handle the algorithmic complexity of the problem. Existing algorithms are now able to efficiently extract patterns with complex structures (ex: sequences, graphs, co-variations) from large datasets. However, when dealing with large, real-world datasets, these methods still output a huge set of patterns, which is impractical for human analysis. This problem is called *pattern explosion*. The ongoing challenge of pattern mining research is to extract fewer
but more meaningful patterns. The LACODAM team is committed to solve the pattern explosion problem by pursuing the following four research topics:

- 1. the design of dedicated algorithms for mining temporal patterns
- 2. the design of flexible pattern mining approaches
- 3. the automatic selection of interesting data mining results
- 4. the design of parallel pattern algorithms to ensure scalability

The originality of our contributions relies on the exploration of knowledge-based approaches whose principle is to incorporate dedicated domain knowledge (aka application background knowledge) deep into the mining process. While most of the data mining approaches are based on agnostic approaches that are designed to cope with the pattern explosion, we propose to develop data mining techniques relying on knowledge-based artificial intelligence techniques. This entails the use of structured knowledge representations, as well as reasoning methods, in combination with mining.

The first topic concerns classical pattern mining in conjunction with expert knowledge in order to define new pattern types (and related algorithms) that can solve applicative issues. In particular, we investigate how to handle temporality in pattern representations which turns out to be important in many real world applications (in particular for decision support) and deserves particular attention.

The next two topics aim at proposing alternative pattern mining methods to let the user incorporate, on her own, knowledge that will help define her pattern domain of interest. Flexible pattern mining approaches enable analysts to easily incorporate extra knowledge, for example domain related constraints, in order to extract only the most relevant patterns. On the other hand, the selection of interesting data mining results aims at devising strategies to filter out the results that are useless for the data analyst. Besides the challenge of algorithmic efficiency, we are interested in formalizing the foundations of interestingness, according to background knowledge modeled with logical knowledge representation paradigms.

Last but not least, pattern mining algorithms are compute-intensive. It is thus important to exploit all the available computing power. Parallelism is for a foreseeable future one of the main ways to speed up computations, and we have a strong competence on the design of parallel pattern mining algorithms. We will exploit this competence in order to guarantee that our approaches scale up to the data provided by our partners.

# 3.3. User/system interaction

As we pointed out before, there is a strong need to present relevant patterns to the user. This can be done by using more specific constraints, background knowledge and/or tailor-made optimization functions. Due to the difficulty of determining these elements beforehand, one of the most promising solutions is that the system and the user co-construct the definition of relevance, i.e., to have a human in the loop. This requires to have means to present intermediate results to the user, and to get user feedback in order to guide the search space exploration process in the right direction. This is an important research axis for LACODAM, which will be tackled in several complementary ways:

- *Domain Specific Languages:* One way to interact with the user is to propose a Domain Specific Language (DSL) tailored to the domain at hand and to the analysis tasks. The challenge is to propose a DSL allowing the users to easily express the required processing workflows, to deploy those workflows for mining on large volumes of data and to offer as much automation as possible.
- *What if / What for scenarios:* We also investigate the use of scenarios to query results from data mining processes, as well as other complex processes such as complex system simulations or model predictions. Such scenarios are answers to questions of the type "what if [situation]?" or "what [should be done] for [expected outcome]?".
- User preferences: In exploratory analysis, users often do not have a precise idea of what they want, and are not able to formulate such queries. Hence, in LACODAM we investigate simple ways for users to express their interests and preferences, either during the mining process to guide the search space exploration –, or afterwards during the filtering and interpretation of the most relevant results.

• Data visualization: Most of the research directions presented in this document require users to examine patterns at some point. The output of most pattern mining algorithms is usually a (long) list of patterns. While this presentation can be sufficient for some applications, often it does not provide a complete understanding, especially for non-experts in pattern mining. A transversal research topic that we want to explore in LACODAM is to propose data visualization techniques that are adequate for understanding output results. Numerous (failed) experiments have shown that data mining and data visualization are fields, which require distinct skills, thus researchers in one field usually do not make significant advances in the other field (this is detailed in [Keim 2010]). Thus, our strategy is to establish collaborations with prominent data visualization teams for this line of research, with a long term goal to recruit a specialist in data visualization if the opportunity arises.

## 3.4. Decision support

Predictive sequential patterns have a direct application in diagnosis. LACODAM inherits a strong background in decision support systems with internationally recognized expertise in diagnosis from the former DREAM team. This AI subfield is concerned with determining whether a system is operating normally or not, and the cause of faulty behaviors. The studied system can be an agro- or eco-system, a software system, a living being, etc.

The increasing volumes of data coming from a range of different systems (ex: sensor data from agroenvironmental systems, log data from software systems, biological data coming from health monitoring systems) can help human and software agents make better decisions. Hence, LACODAM builds upon the idea that decision support systems (an interest bequeathed from DREAM) should take advantage of the available data. This third and last research axis is thus a meeting point for all members of the team, as it requires the integration of AI techniques for traditional decision support systems with results from data mining techniques.

Two main research sub-axes are investigated in LACODAM:

- *Diagnosis-based approaches.* We are exploring how to integrate knowledge found from pattern mining approaches, possibly with the help of interactive methods, into the qualitative models. The goal of such work is to automate as much as possible the construction of prediction models, which can require a lot of human effort.
- Actionable patterns and rules. In many settings of "exploratory data mining", the actual interestingness of a pattern is hard to assess, as it may be subjective. However, for some applications there are well defined measures of interestingness and applicability for patterns. Patterns and rules that can lead to actual actions –that are relevant to the user– are called "actionable patterns" and are of vital importance to industrial settings.

# 3.5. Long-term goals

The following perspectives are at the convergence of the three aforementioned research axes and can be seen as ideal towards our goals:

• Automating data science workflow discovery. The current methods for knowledge extraction and construction of decision support systems require a lot of human effort. Our three research axes aim at alleviating this effort, by devising methods that are more generic and by improving the interaction between the user and the system. An ideal solution would be that the user could forget completely about the existence of pattern mining or decision support methods. Instead the user would only loosely specify her problem, while the system constructs various data science / decision support workflows, possibly further refined via interactions.

We consider that this is a second order AI task, where AI techniques such as planning are used to explore the workflow search space, the workflow itself being composed of data mining and/or decision support components. This is a strategic evolution for data science endeavors, were the demand far exceeds the available human skilled manpower. • Logic argumentation based on epistemic interest. Having increasingly automated approaches will require better and better ways to handle the interactions with the user. Our second long term goal is to explore the use of logic argumentation, i.e., the formalisation of human strategies for reasoning and arguing, in the interaction between users and data analysis tools. Alongside visualization and interactive data mining tools, logic argumentation can be a way for users to query both the results and the way they are obtained. Such querying can also help the expert to reformulate her query in an interactive analysis setting.

This research direction aims at exploiting principles of interactive data analysis in the context of epistemic interestingness measures. Logic argumentation can be a natural tool for interactions between the user and the system: display of possibly exhaustive list of arguments, relationships between arguments (e.g., reinforcement, compatibility or conflict), possible solutions for argument conflicts, etc.

The first step is to define a formal argumentation framework for explaining data mining results. This implies to continue theoretical work on the foundations of argumentation in order to identify the most adapted framework (either existing or a new one to be defined). Logic argumentation may be implemented and deeply explored in ASP, allowing us to build on our expertise in this logic language.

• *Collaborative feedback and knowledge management.* We are convinced that improving the data science process, and possibly automating it, will rely on high-quality feedback from communities on the web. Consider for example what has been achieved by collaborative platforms such as StackOverflow: it has become the reference site for any programming question.

Data science is a more complex problem than programming, as in order to get help from the community, the user has to share her data and workflow, or at least some parts of them. This raises obvious privacy issues that may prevent this idea to succeed. As our research on automating the production of data science workflows should enable more people to have access to data science results, we are interested in the design of collaborative platforms to exchange expert advices over data, workflows and analysis results. This aims at exploiting human feedback to improve the automation of data science system via machine learning methods.

# 4. Application Domains

# 4.1. Introduction

The current period is extremely favorable for teams working in Data Science and Artificial Intelligence, and LACODAM is not the exception. We are eager to see our work applied in real world applications, and have thus an important activity in maintaining strong ties with industrials partners concerned with marketing and energy as well as public partners working on health, agriculture and environment.

# 4.2. Industry

We present below our industrial collaborations. Some are well established partnerships, while others are more recent collaborations with local industries that wish to reinforce their Data Science R&D with us (e.g. STMicroelectronics, Energiency, Amossys).

• Execution trace analysis for SOC debugging (STMicroelectronics). We have an ongoing collaborations with STMicroelectronics, which is one of top-5 electronic chip makers worldwide. Nowadays, set-top boxes, smartphones and onboard car computers are powered by highly integrated chips called System-on-Chip (SoC). Such chips contain on a single die, processing units, memories, IO units and specialized accelerators (such as audio and video encoding/decoding). Programming SoC is a hard task due to their inherent parallelism, leading to subtle bugs when several components do not deliver their results within a given time frame. Existing debuggers and profilers are ill-adapted in this case because of their high intrusivity that modifies the timings. Hence the most used technique is to capture a trace of the execution and analyze it post-mortem. While Alexandre Termier was in Grenoble he initiated several works for analyzing such traces with pattern mining techniques, which he is now pursuing with his colleagues of the LACODAM project-team.

- Resource consumption analysis for optimizing energy consumption and practices in industrial factories (Energiency). In order to increase their benefits, companies introduce more and more sensors in their factories. Thus, the resource (electricity, water, etc.) consumption of engines, workshops and factories are recorded in the form of times series or temporal sequences. The person who is in charge of resource consumption optimization needs better software than classical spreadsheets for this purpose. He/she needs effective decision-aiding tools with statistical and artificial intelligence knowledge. The start-up Energiency aims at designing and offering such pieces of software for analyzing energy consumption. The starting CIFRE PhD thesis of Maël Guillemé aims at proposing new approaches and solutions from the data mining field to tackle this issue.
- Security (Amossys). Current networks are faced with an increasing variety of attacks, from the classic "DDoS" that makes a server unusuable for a few fours, to advanced attacks that silently infiltrate a network and exfiltrate sensitive information months or even years later. Such intrusions, called APT (Advanced Persistent Threat) are extremely hard to detect, and this will become even harder as most communications will be encrypted. A promising solution is to work on "behavioral analysis", by discovering patterns based on the metadata of IP-packets. Such patterns can relate to an unusual sequencing of events, or to an unusual communication graph. Finding such complex patterns over a large volume of streaming data requires to revisit existing stream mining algorithms to dramatically improve their throughput, while guaranteeing a manageable false positive rate. We are collaborating on this topic with the Amossys company and the EMSEC team of Irisa through the co-supervision of a CIFRE PhD (located in the EMSEC team). Our goal is to design novel anomaly detection methods that can detect APT, and that scales on real traffic volumes.
- Market basket data analysis (Intermarché) and multi-channel interaction data analysis (EDF) for better Customer Relationship Management (CRM). An important application domain of data mining for companies that deal with large numbers of customers is to analyze customer interaction data, either for marketing purposes or to improve the quality of service. We have activities in both settings. In the first case, we collaborate with a major french retailer, Intermarché, in order to detect customer churn by analyzing market basket data. In the second case, we collaborate with the major french power supplier, EDF, to discover actionable patterns for CRM that aim at avoiding undesirable situations. We use logs of user interactions with the company (e.g., web clicks, phone calls, etc.) for this purpose.

# 4.3. Health

• Care pathways analysis for supporting pharmaco-epidemiological studies. Pharmaco-epidemiology applies the methodologies developed in general epidemiology to answer to questions about the uses and effects of health products, drugs [33], [32] or medical devices [27], on population. In classical pharmaco-epidemiology studies, people who share common characteristics are recruited to build a dedicated prospective cohort. Then, meaningful data (drug exposures, diseases, etc.) are collected from the cohort within a defined period of time. Finally, a statistical analysis highlights the links (or the lack of links) between drug exposures and outcomes (*e.g.* adverse effects). The main drawback of prospective cohort studies is the time required to collect the data and to integrate them. Indeed, in some cases of health product safety, health authorities have to answer quickly to pharmaco-epidemiology questions.

New approaches of pharmaco-epidemiology consist in using large EHR (Electronic Health Records) databases to investigate the effects and uses (or misuses) of drugs in real conditions. The objective is to benefit from nationwide available data to answer accurately and in a short time pharmaco-epidemiological queries for national public health institutions. Despite the potential availability of

the data, their size and complexity make their analysis long and tremendous. The challenge we tackle is the conception of a generic digital toolbox to support the efficient design of a broad range of pharmaco-epidemiology studies from EHR databases.

We propose to use pattern mining algorithm and reasoning techniques to analyse the typical care pathways of specific groups of patients.

To answer the broad range of pharmaco-epidemiological queries from national public health institutions, the PEPS <sup>0</sup> platform exploits, in secondary use, the French health cross-schemes insurance system, called SNDS. The SNDS covers most of the French population with a sliding period of 3 past years. The main characteristics of this data warehouse are described in [31]. Contrary to local hospital EHR or even to other national initiatives, the SNDS data warehouse covers a huge population. It makes possible studies on unfrequent drugs or diseases in real conditions of use. To tackle the volume and the diversity of the SNDS data warehouse, a research program has been established to design an innovative toolbox. This research program is focused first on the modeling of care pathways from the SNDS database and, second, on the design of tools supporting the extraction of insights about massive and complex care pathways by clinicians. In such a database a care pathway is an individual sequence of drugs exposures, medical procedures and hospitalizations.

## 4.4. Agriculture and environment

• Dairy farming. The use and analysis of data acquired in dairy farming is a challenge both for data science and animal science. The goal is to improve farming conditions, i.e., health, welfare and environment, as well as farmers' income. Nowadays, animals are monitored by multiple sensors giving a wealth of heterogeneous data, ex: temperature, weight, or milk composition. Current techniques used by animal scientists focus mostly on mono-sensor approaches. The dynamic combination of several sensors could provide new services and information useful for dairy farming. The PhD thesis of Kevin Fauvel (#DigitAg grant), aims to study such combinations of sensors and to investigate the use data mining methods, especially pattern mining algorithms. The challenge is to design new algorithms that take into account the data heterogeneity –in terms of nature and time units–, and that produce useful patterns for dairy farming. The outcome of this thesis will be an original and important contribution to the new challenge of the IoT (Internet of Things) and will interest domain actors to find new added value to a global data analysis. The PhD thesis, started on October 2017, takes place in an interdisciplinary setting bringing together computer scientists from Inria and animal scientists from INRA, both located in Rennes.

Similar problems are investigated with the veterinary department of the University of Calgary in the context of cattle monitoring from multiple sensors placed on calves for the early detection of diseases.

- **Optimizing the nutrition of individual sow**. Another direction for further research is the combination of data flows with prediction models in order to learn nutrition strategies. Raphaël Gauthier started a PhD thesis (#DigitAg Grant) in November 2017 with both Inria and INRA supervisors. His research addresses the problem of finding the optimal diet to be supplied to individual sows. Given all the information available, e.g., time-series information about previous feeding, environmental data, scientists models, the research goal is to design new algorithms to determine the optimal ration for a given sow in a given day. Efficiency issues of developed algorithms will be considered since the proposed software should work in real-time on the automated feeder. The decision support process should involve the stakeholder to ensure a good level of acceptance, confidence and understanding of the final tool.
- Ecosystem modeling and management. Ongoing research on ecosystem management includes modelling of ecosystems and anthroprogenic pressures, with a special concern on the representation of socio-economical factors that impact human decisions. A main research issue is how to represent

<sup>&</sup>lt;sup>0</sup>PEPS: Pharmaco-Epidémiologie et Produits de Santé – Pharmacoepidemiology of health products

these factors and how to integrate their impact on the ecosystem simulation model. This work is an ongoing cooperation with ecologists from the Marine Spatial Ecology of Queensland University, Australia and from Agrocampus Ouest.

# 5. Highlights of the Year

# 5.1. Highlights of the Year

This year's highlight was the exceptional success of LACODAM in the recruiting process: we could hire two new staff members.

- Elisa Fromont joined as a Professor of University of Rennes 1. She brings to the team her skills in Machine Learning, which include a precious expertise on dealing with numerical data.
- Luis Galárraga Del Prado joined as an Inria Researcher. He brings to the team his skills in exploiting knowledge bases, which will be strongly needed for including domain knowledge in our approaches. He will also reinforce our work on rule mining.

# 6. New Software and Platforms

## 6.1. EcoMata

KEYWORD: Environment perception

FUNCTIONAL DESCRIPTION: The EcoMata toolbox provides means for qualitative modeling and exploration of ecosystems in order to aid the design of environmental guidelines. We have proposed a new qualitative approach for ecosystem modeling based on the timed automata (TA) formalism combined to a high-level query language for exploring scenarios.

- Participants: Christine Largouët, Marie-Odile Cordier, Thomas Guyet and Yulong Zhao
- Contact: Christine Largouët
- URL: https://team.inria.fr/dream/fr/ecomata/

# **6.2. PATURMATA**

**KEYWORDS: Bioinformatics - Biology** 

SCIENTIFIC DESCRIPTION: The Paturmata tool-box provides means for qualitative modeling and the exploration of agrosystems, specifically management of herd based on pasture. The system is modeled using a hierarchical hybrid model described in the timed automata formalism.

FUNCTIONAL DESCRIPTION: In the PaturMata software, users can create a pasture system description by entering herds and plots information. For each herd, the only parameter is the number of animals. For each plot, users should enter the surface, the density, the herb height, the distance to the milking shed, a herb growth profile and an accessibility degree. Users then specify pasturing and fertilization strategies. Finally, users can launch a pasture execution. PaturMata displays the results and a detailed trace of pasture. Users can launch a batch of different strategies and compare the results in order to find the best pasture strategy. PaturMata is developed in Java (Swing for the GUI) and the model-checker that is called for the timed properties verification is UPPAAL.

- Participants: Christine Largouët and Marie-Odile Cordier
- Contact: Christine Largouët

# 6.3. Promise

**KEYWORDS:** Data mining - Monitoring

FUNCTIONAL DESCRIPTION: Promise is a software that predicts rare events in industrial production systems from data analysis of energy consumption data. The data is represented as a time series. The program takes as input the temporal series of energy consumption, an abnormal pattern (rare event) and a temporal dilatation, and outputs a set of sub-series similar (according to a similarity metric) to the abnormal pattern.

- Participants: Véronique Masson, Laurence Rozé and Mael Guilleme
- Contact: Véronique Masson

# 6.4. GWASDM

## Genome Wide Association Study using Data Mining strategy

KEYWORDS: GWAS - Data mining

FUNCTIONAL DESCRIPTION: From two cohorts of genotyped individuals (case and control), the GWASDM software performs a Genome Wide Association Study based on data mining techniques and generates several patterns of SNPs that correlate with a given phenotype. The algorithm implemented in GWASDM directly uses relative risk measures such as risk ratio, odds ratio and absolute risk reduction combined with confidence intervals as anti-monotonic properties to efficiently prune the search space. The algorithm discovers a complete set of discriminating patterns with regard to given thresholds or applies heuristic strategies to extract the largest statistically significant discriminating patterns in a given dataset.

• Contact: Dominique Lavenier

# 6.5. DCM

#### Discriminant Chronicle Mining

KEYWORDS: Pattern extraction - Sequence - Classification

FUNCTIONAL DESCRIPTION: DCM is a temporal sequences analysis tool. It extracts discriminant chronicles from a large set of labeled sequences. A sequence is made of timestamped events. Each sequence of events is associated to a label (e.g. positive and negative sequences). A chronicle is a temporal model that characterizes a behavior by a set of events linked by temporal constraints. The DCM algorithm extracts chronicles that occurs more in positive sequences than in negative sequences.

- Participants: Yann Dauxais and Thomas Guyet
- Partners: REPERES Université de Rennes 1
- Contact: Yann Dauxais
- Publications: Discriminant chronicles mining: Application to care pathways analytics Extraction de chroniques discriminantes
- URL: https://gitlab.inria.fr/ydauxais/DCM

# 6.6. NTGSP

#### Negative Time-Gap Sequential Patterns

KEYWORDS: Pattern discovery - Sequence

FUNCTIONAL DESCRIPTION: The NTGSP algorithm is a sequential pattern mining algorithm. It analyses a large database of temporal sequences, i.e., events with timestamps, by extracting its regularities (the patterns). A pattern describes the behavior as a sequence of events that frequently occurred in sequences. What makes NTGSP novel is its ability to handle patterns with negations, i.e., the description of a behavior that specifies the absence of an event. More precisely, it extracts frequent sequences with positive and negative events, as well as temporal information about the delay between these events.

- Participants: Thomas Guyet and René Quiniou
- Partner: Edf
- Contact: René Quiniou
- Publication: Fouille de motifs temporels négatifs

# 6.7. Relevant Interval Rules Miner

KEYWORDS: Association rule - Pattern discovery - Formal concept analysis

FUNCTIONAL DESCRIPTION: This software extracts relevant rules from a dataset of labeled numerical attributes (tabular datasets). A rule is an interval-based pattern associated to a predicted label. The tool extracts a subset of rules based on the accuracy and relevance criteria where most of the algorithms are simply based on accuracy. This allow us to extract the best rules that capture the data behavior.

- Participants: René Quiniou, Véronique Masson and Thomas Guyet
- Contact: Thomas Guyet
- Publication: Mining relevant interval rules

# 6.8. OCL

#### One click learning

KEYWORDS: Data mining - Interactivity

FUNCTIONAL DESCRIPTION: This pattern mining software builds a user model preference from implicit feedback of the user in order to automatically choice the type of patterns and algorithms used. The principle builds upon the algorithm introduced by M. Boley et al, "One click mining: interactive local pattern discovery through implicit preference and performance learning". In addition OCL integrates algorithms dealing with temporal series.

- Contact: Laurence Rozé
- URL: https://github.com/Gremarti/OneClickLearning

# 7. New Results

# 7.1. Introduction

In this section, we organize our contributions this year along two of our research axes, namely Pattern Mining and Decision Support. These correspond to the contributions that has been accepted for publication this year.

#### 7.1.1. Pattern Mining

In the domain of pattern mining we can categorize our contributions along the following lines:

- *Mining of novel types of patterns*. This includes temporal pattern mining, signature mining, opinion mining in uncertain databases, interval rules, and top-k item-centric mining. All these contributions have been proposed as solutions to problems in the domains of pharmaco-epidemiology, retail databases, biomedical databases, and analysis of speech corpora. We provide more details about these results in Sections 7.2 to 7.9.
- *Data Mining with ASP.* Answer Set Programming is a powerful search tool in combinatorial spaces, which can be naturally ported to pattern mining, as the latter is a specific type of search problem. Our contributions include the application of ASP in the discovery of frequent, constrained, condensed, and rare sequential patterns. Sections 7.11 and 7.12 elaborate on our new research insights.
- Data Mining for the masses. In [14], we propose a communication model that bridges knowledge delivery between data miners and domain users in the field of library science. Our model proposes a five-steps process in order to achieve effective knowledge synthesis and delivery of insights to the domain users.

#### 7.1.2. Decision Support

In regards to the axis of decision support, our contributions can be organized in two categories: exploration and diagnosis.

- *Exploration.* We propose two exploration methods in the context of analysis of trajectories and agro-environmental systems. We propose customized data models and resort to data-warehousing and multidimensional data representations to facilitate the querying, and thus the exploration and understanding of the data, for the sake of decision making. Our results in this line are further detailed in Sections 7.13 to 7.15.
- *Diagnosis.* In Section 7.16 we propose a novel method for anomaly detection in time series by resorting to Extreme Value Theory. In addition, [21] offers a formalization of diagnosis based on automata with focus on discrete event systems.

#### 7.2. Discriminant chronicles mining: Application to care pathways analytics

**Participants:** Yann Dauxais, Thomas Guyet, David Gross-Amblard [Druid], André Happe [Brest University Hospital/REPERES].

Pharmaco-epidemiology (PE) is the study of uses and effects of drugs in well defined populations. As medicoadministrative databases cover a large part of the population, they have become very interesting to carry PE studies. Such databases provide longitudinal care pathways in real condition containing timestamped care events, especially drug deliveries. Temporal pattern mining becomes a strategic choice to gain valuable insights about drug uses. We propose DCM [8], [7], a new discriminant temporal pattern mining algorithm. It extracts chronicle patterns that occur more in a studied population than in a control population. We present satisfactory results on the identification of possible associations between hospitalizations for seizure and anti-epileptic drug switches in care pathway of epileptic patients.

A stable release of the DCM algorithm (see Section 6.5) have been deposed to the Program Protection Agency (APP) and is available online.

## 7.3. Purchase Signatures of Retail Customers

Participants: Clément Gautrais, Peggy Cellier [SemLis], Thomas Guyet, René Quiniou, Alexandre Termier.

In the retail context, there is an increasing need for understanding individual customer behavior in order to personalize marketing actions. We propose the novel concept of customer signature, that identifies a set of important products that the customer refills regularly [10]. Both the set of products and the refilling time periods give new insights on the customer behavior. Our approach is inspired by methods from the domain of sequence segmentation, thus benefiting from efficient exact and approximate algorithms. Experiments on a real massive retail dataset show the applicability of the signatures for understanding individual customers.

# 7.4. Topic Signatures in Political Campaign Speeches

Participants: Clément Gautrais, Peggy Cellier [SemLis], René Quiniou, Alexandre Termier.

Highlighting the recurrence of topics usage in candidates speeches is a key feature to identify the main ideas of each candidate during a political campaign. In this study [9], we develop a method combining standard topic modeling with signature mining for analyzing topic recurrence in speeches of Clinton and Trump during the 2016 American presidential campaign. The results show that the method extracts automatically the main ideas of each candidate and, in addition, provides information about the evolution of these topics during the campaign.

# 7.5. Expert Opinion Extraction from a Biomedical Database

**Participants:** Ahmed Samet, Thomas Guyet, Benjamin Négrevergne, Tien-Tuan Dao, Tuan Nha Hoang, Marie-Christine Ho Ba Tho.

This work tackles the problem of extracting frequent opinions from uncertain databases. This problem is encountered in real-world applications, such as the opinions of medical experts to evaluate the reliability level of biomedical data. We introduce the foundation of an opinion mining approach with the definition of pattern and support measure. The support measure is derived from the commitment definition. In [15], we proposed a new algorithm called OPMINER that extracts the set of frequent opinions modeled as a mass functions. We applied it on a real-world biomedical opinion database. Performance analysis showed that our proposal generated better patterns compared to literature-based methods.

## 7.6. Mining Relevant Interval Rules

Participants: Philippe Besnard, Thomas Guyet, Véronique Masson, René Quiniou.

Rule mining is a classical data mining task. Numerical rule mining consists of extracting decision rules from a dataset with numerical attributes. In this work, we are interested in extracting a subset of accurate rules, called relevant rules. This selection criteria was introduced by Garriga et al. for categorical attributes [28]. In [13] we extend the method of Garriga et *al.* for mining relevant rules on numerical attributes by extracting interval-based pattern rules. We proposed an algorithm that extracts such rules from numerical datasets using the interval-pattern approach from Kaytoue et *al.* [29]. The algorithm has been implemented and intensively evaluated on real datasets. This study on numerical rules mining leads us to initiate a study about admissible generatizations of examples as rules [18].

## 7.7. Time Series Rule Matching: Application to Energy Consumption

Participants: Maël Guillemé, Véronique Masson, Laurence Rozé, René Quiniou, Alexandre Termier.

Pattern mining in time series is an important subfield of Data Mining. In various applications, patterns exhibit distortion in time (or time elasticity) that requires using specific distance measures. In this work, we extend an algorithm proposed by Shokoohi et *al.* [35] by improving the performance of rule matching in the detection of energy consumption patterns. Nowadays companies are more and more equipped with sensors in order to trace losses of energy resources. Detecting dysfunctions from time series recorded by these sensors becomes a crucial problem for reducing energy consumption. Locating specific patterns related to dysfunctions in time series requires handling with time elasticity (i.e. distortion in time) of patterns. We propose a detection of predictive rules based on several variations of Dynamic Time Warping (DTW) and show the superiority of subsequence DTW [11]. We study now multivariate time series classification to predict dysfunctions as soon as possible.

# 7.8. Negative Temporal Sequence Mining

**Participants:** Katerina Tsesmeli, Thomas Guyet, René Quiniou, Manel Boumghar [EDF R&D], Laurent Pierre [EDF R&D].

Temporal pattern mining is one of the important tasks in the data mining research field. It aims at extracting interesting sequences of occurring events from timestamped event sequences as well as their temporal constraints relating sequence events. Little research has focused on mining sequential patterns with non-occurring (negative) events, though they can bring much value and relevance to extracted patterns. In this context, we are interested in formalizing normal and undesirable situations, that can be defined in terms of negative temporal patterns. We proposed the NTGSP algorithm [17] that extracts frequent sequences with positive and negative events, as well as temporal information about the delay between these events. The method performance has been evaluated on synthetic sequences and on commercial data provided by EDF, a major french power distribution company.

# 7.9. TopPI: An efficient algorithm for item-centric mining

Participants: Vincent Leroy, Martin Kirchgessner, Alexandre Termier, Sihem Amer-Yahia.

In this paper [6], we introduce item-centric mining, a new semantics for mining long-tailed datasets. Our algorithm, TopPI, finds for each item its top-k most frequent closed itemsets. While most mining algorithms focus on the globally most frequent itemsets, TopPI guarantees that each item is represented in the results, regardless of its frequency in the database.

TopPI allows users to efficiently explore Web data, answering questions such as "what are the k most common sets of songs downloaded together with the ones of my favorite artist?". When processing retail data consisting of 55 million supermarket receipts, TopPI finds the itemset "milk, puff pastry" that appears 10,315 times, but also "frangipane, puff pastry" and "nori seaweed, wasabi, sushi rice" that occur only 1120 and 163 times, respectively. Our experiments with analysts from the marketing department of our retail partner, demonstrate that item-centric mining discover valuable itemsets. We also show that TopPI can serve as a building-block to approximate complex itemset ranking measures such as the p-value.

Thanks to efficient enumeration and pruning strategies, TopPI avoids the search space explosion induced by mining low support itemsets. We show how TopPI can be parallelized on multi-core architectures and Hadoop clusters. Our experiments on datasets with different characteristics show the superiority of TopPI when compared to standard top-k solutions, and to Parallel FPGrowth, its closest competitor.

#### 7.10. Declarative Sequential Pattern Mining of Care Pathways

Participants: Thomas Guyet, André Happe [Brest University Hospital/REPERES], Yann Dauxais.

Sequential pattern mining algorithms are widely used to explore care pathways database, but they generate a deluge of patterns, mostly redundant or non-informative. Clinicians need tools to express complex mining queries in order to generate less but more significant patterns. These algorithms are not versatile enough to answer complex clinician queries. This work [12] proposes to apply a declarative pattern mining approach based on the Answer Set Programming paradigm. It is exemplified by a pharmaco-epidemiological study investigating the possible association between hospitalization for seizure and antiepileptic drug switch from a French medico-administrative database.

# 7.11. Efficiency Analysis of ASP Encodings for Sequential Pattern Mining Tasks

Participants: Thomas Guyet, Yves Moinard, René Quiniou, Torsten Schaub.

This study [22] presents the use of Answer Set Programming (ASP) to mine sequential patterns. ASP is a highlevel declarative logic programming paradigm that allows for representation of combinatorial and optimization problems, as well as knowledge and reasoning tasks. Thus, ASP is a good candidate for implementing pattern mining with background knowledge, which has been a data mining issue for a long time. We propose encodings of the classical sequential pattern mining tasks within two representations of embeddings (fill-gaps vs skipgaps) and for various kinds of patterns: frequent, constrained and condensed. We compare the computational performance of these encodings with each other to get a good insight into the efficiency of ASP encodings. The results show that the fill-gaps strategy is better on real problems due to lower memory consumption. Finally, compared to a constraint programming approach (CPSM), another declarative programming paradigm, our proposal showed comparable performance.

# 7.12. Mining Rare Sequential Patterns with ASP

Participants: Ahmed Samet, Thomas Guyet, Benjamin Négrevergne.

This work [20] presents an approach of meaningful rare sequential pattern mining based on the declarative programming paradigm of Answer Set Programming (ASP). The setting of rare sequential pattern mining is introduced. Our ASP approach provides an easy manner to encode expert constraints on expected patterns to cope with the huge amount of meaningless rare patterns. Encodings are presented and quantitatively compared to a procedural baseline. An application on care pathways analysis illustrates the applicability of our method in the encoding of constraints provided by experts.

# 7.13. From Medico-administrative Databases Analysis to Care Trajectories Analytics: An Example with the French SNDS

**Participants:** Erwan Drezen [Rennes University Hospital/REPERES], Thomas Guyet, André Happe [Brest University Hospital/REPERES].

Medico-administrative data like SNDS (Système National de Données de Santé) are not collected initially for epidemiological purposes. Moreover, the data model and the tools proposed to SNDS users make their in-depth exploitation difficult. We propose a data model, called the ePEPS model, based on health care trajectories to provide a medical view of raw data [4]. A data abstraction process enables the clinician to have an intuitive medical view of raw data and to design a study-specific views. This view is based on a generic model of care trajectory, i.e. a sequence of timestamped medical events for a given patient. This model is combined with tools to manipulate care trajectories efficiently.

# 7.14. A Data Warehouse to Explore Multidimensional Simulated Data from a Spatially Distributed Agro-hydrological Model to Improve Catchment Nitrogen Management

**Participants:** Tassadit Bouadi, Marie-Odile Cordier, Pierre Moreau, Jordy Salmon-Monviola, Chantal Gascuel-Odoux.

Spatially distributed agro-hydrological models allow researchers and stakeholders to represent, understand and formulate hypotheses about the functioning of agro-environmental systems and to predict their evolution. These models have guided agricultural management by simulating effects of landscape structure, farming system changes and their spatial arrangement on stream water quality. Such models generate many intermediate results that should be managed, analyzed and transformed into usable information. We introduce [3] a data warehouse (N-Catch) built to store and analyze simulation data from the spatially distributed agro-hydrological model TNT2. We present scientific challenges to and tools for building data warehouses and describe the three dimensions of N-Catch: space, time and an original hierarchical description of cropping systems. We show how to use OLAP to explore and extract all kinds of useful high-level information by aggregating the data along these three dimensions. We also show how to facilitate exploration of the spatial dimension by coupling N-Catch with GIS. Such tool constitutes an efficient interface between science and society, simulation remaining a research activity, exploration of the results becoming an easy task accessible for a large audience.

# 7.15. Extended Automata for Temporal Planning of Interacting Agents

Participants: Christine Largouët, Omar Krichen, Yulong Zhao.

In this paper [5], we consider the planning problem for a system represented as a set of interacting agents evolving along time according to explicit timing constraints. Given a goal, the planning problem is to find the sequence of actions such that the system reaches the goal state in a limited time and in an optimal manner, assuming actions have a cost. In our approach, the planning problem is based on model-checking and controller synthesis techniques while the goal is defined using temporal logic. Each agent of the system is represented using the formalism of Priced Timed Game Automata (PTGA). PTGA is an extension of Timed Automata that allows the representation of cost on actions and the definition of uncontrollable actions. We define a planning algorithm that computes the best strategy to achieve a goal. To experiment our approach, we extend the classical Transport Domain with timing constraints, cost on actions and uncontrollable actions. The planning algorithm is finally presented on a marine ecosystem management problem.

# 7.16. Anomaly Detection in Streams with Extreme Value Theory

Participants: Alban Siffer [EMSEC], Pierre-Alain Fouque [EMSEC], Christine Largouët, Alexandre Termier.

Anomaly detection in time series has attracted considerable attention due to its importance in many real-world applications including intrusion detection, energy management and finance. Most approaches for detecting outliers rely on either manually set thresholds or assumptions on the distribution of data. In [16], we propose a new approach to detect outliers in streaming univariate time series based on Extreme Value Theory that does not require to handpick thresholds and makes no assumption on the distribution: the main parameter is only the risk, controlling the number of false positives. Our approach can be used for outlier detection, but more generally for automatically setting thresholds, making it useful in wide number of situations. We also test our algorithms on various real-world datasets which confirm the soundness and efficiency of our methods.

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Contracts with Industry

#### 8.1.1. ITRAMI: Interactive Trace Mining

Participant: Alexandre Termier.

ITRAMI is a Nano2017 project. Such projects are designed to support joint research efforts between STMicrolectronics and academic partners in the domain of embedded systems. Alexandre Termier is the PI of this project whose goal is to design novel data mining methods for interactive analysis of execution traces. Such methods aim at considerably reducing the time that STMicroelectronics developers spend at understanding, debugging and profiling applications running on STMicrolectronics chips. The work is done at University Grenoble Alps, in collaboration with LACODAM researchers. Two contractual staff members are working on the project in Grenoble: Willy Ugarte as postdoc, and Soumaya Ben Alouane as engineer.

#### 8.1.2. Hyptser: Hybrid Prediction of Time-Series

Participants: Thomas Guyet, Vincent Lemaire [Orange Labs], Simon Malinowski [LinkMedia].

HYPTSER is a project funded by the Gaspard Monge Program for Optimisation and Operational Research (PGMO). It is dedicated to the development of innovative methods for predictions in time series. In the field of machine learning, *ensemble methods* have gained popularity in the last years. These methods combine several algorithms that solve the same task in order to improve the performance of the outcome. Two main families of ensemble methods can be found in the literature : The first family makes use of different models and combine their results a posteriori. The methods Bagging and Boosting are examples of methods in this family [26], [34]. The second family is based on a smart selection of the local algorithms in order to create a global hybrid algorithm. Logistic Model Tree [30] or Extreme Learning Machine Tree [36] are examples of such hybrid algorithms. In this project, starting at the end of 2017 for one year, we envision to explore the second family of methods in order to analyze how efficiently hybrid models can perform on the task of time series prediction. We plan to apply these methods to predict resource usage for cloud computing (CPU, memory) so as to minimize their infrastructure.

#### 8.1.3. Particular Contract of the Strategic Action EDF/Inria

Participants: Manel Boumghar [EDF R&D], Laurent Pierre [EDF R&D], Thomas Guyet, René Quiniou.

The analysis of customer pathways has become a strategic issue for many businesses. The interaction traces left by clients when connecting to the customer services can be combined with data from other communication channels (phone, web form, e-mail, mail, fax, SMS, shop, etc.) and allow to analyse the customer pathways in details.

Pattern mining tools are able to extract the frequent customer behaviors in very large databases of client pathways. Nevertheless, taking into account the duration and the delay between the customer actions in the mining remains a challenge. The objective of this one-year contract was to design and develop a frequent mining tool that accounts for temporal patterns with negations for analysis of multichannel customer pathways. In this line, we developed and implemented the NTGSP algorithm [17].

# 8.2. Bilateral Grants with Industry

Maël Guillemé has obtained a CIFRE PhD grant with the Energiency startup, supervised by V. Masson and L. Rozé. The goal of Maël's thesis is to propose new approaches to improve industrial energy performance by integrating both numerical and symbolic attributes. An M2 internship from 2016 explored an approach based on an algorithm proposed by Shokoohi and al, and proposed several improvements: avoid data normalisation, detect patterns as fast as possible, enhance functions like distance and score.

Another CIFRE thesis has started, this time with the Amossys company, which specializes in cyber-security. This is the PhD of Alban Siffer, located in the EMSEC team of IRISA and co-supervised between EMSEC (P.A. Fouque) and LACODAM (A. Termier, C. Largouët). The goal of this PhD is to propose new methods for intrusion detection in networks. The novel insight is to consider only IP flow as input (metadata of packets and not packet contents) and detect intrusion via unusual traffic patterns.

On October 2017, Colin Leverger started a thesis funded by Orange and co-supervised between Orange Labs (R. Marguerie), LACODAM (A. Termier, T. Guyet) and LinkMedia (S. Malinowski). The goal of this thesis is to propose new methods to forecast time series in order to support capacity planning tasks.

Elisa Fromont is still involved in the supervision of two PhD students through her former employer: the University of Saint-Etienne. One of the students is Guillaume Metzler, who works with the sponsorship of the Blitz company on bank fraud detection. On the other hand, Kevin Bascol (financed by a FUI project) works in collaboration with Bluecime (Grenoble) and works on improving ski-lift security.

# 9. Partnerships and Cooperations

# 9.1. Regional Initiatives

#### 9.1.1. SePaDec: Declarative approaches for Sequential Pattern mining

Participants: Benjamin Negrevergne, Thomas Guyet, Ahmed Samet, Alexandre Termier.

The SEPADEC project is funded by the Region Bretagne. During the execution of this project we explored the application of declarative pattern mining (specifically ASP) in the field of care pathway analysis. The goal was to model domain knowledge to enrich raw data with medical expert knowledge and to develop a toolbox that smoothly integrates both expert knowledge and declarative pattern mining.

We developed a new approach for mining rare sequential mining with ASP [20] and we also proposed a general framework based on ASP for flexibly mine care pathways [12].

# 9.2. National Initiatives

## 9.2.1. ANR

9.2.1.1. #DigitAg: Digital agriculture

Participants: Alexandre Termier, Véronique Masson, Christine Largouët, Anne-Isabelle Graux.

#DigitAg is a "Convergence Institute" dedicated to the increasing importance of digital techniques in agriculture. Its goal is twofold: First, make innovative research on the use of digital techniques in agriculture in order to improve competitiveness, preserve the environment, and offer correct living conditions to farmers. Second, prepare future farmers and agricultural policy makers to successfully exploit such technologies.

While #DigitAg is based on Montpellier, Rennes is a satellite of the institute focused on cattle farming. LACODAM is involved in the "data mining" challenge of the institute, that A. Termier co-leads. He is also the representative of Inria in the steering comittee of the institute.

The interest for the team is to design novel methods to analyze and represent agricultural data, which are challenging because they are both heterogeneous and multi-scale (both spatial and temporal).

#### 9.2.2. National Platforms

#### 9.2.2.1. PEPS: Pharmaco-epidemiology for Health Products

Participants: Yann Dauxais, Thomas Guyet, Véronique Masson, René Quiniou, Ahmed Samet.

The PEPS project (Pharmaco-epidemiology des Produits de Santé) is funded by the ANSM (National Agency for Health Security). The project leader is E. Oger from the clinical investigation center CIC-1414 INSERM/CHU Rennes. The other partners located in Rennes are the Institute of Research and Technology (IRT), B<>Com, EHESP and the LTSI. The project started in January 2015 and is funded for 4 years.

The PEPS project consists of two parts: a set of clinical studies and a research program dedicated to the development of innovative tools for pharmaco-epidemiological studies with medico-administrative databases.

Our contribution to this project will be to propose pattern mining algorithms and reasoning techniques to analyse the typical care pathways of specific groups of insured patients. This year we worked on the design and development of the DCM algorithm [8], [7] to mine patterns on care pathways.

# 9.3. International Research Visitors

#### 9.3.1. Internships

This year, we hosted Scarlett Kelly, a student of Dalhousie University (Canada) from May to the end of August. Her internship was funded by a joint Mitacs Globalink (Canada) / Inria grant. Scarlett Kelly is a student of social sciences, thus she has a different profile that the computer science students who usually do internships at LACODAM. We were interested in such profile in order to gain a critical view on the current approaches of *interactive data mining*. Scarlett quickly picked up the literature of the domain, and could write a report and make interesting propositions that were unexpected from a computer science point of view, i.e., introduce a specially trained "data liaison" person between practitioners and data scientists. Her proposition led to a paper [14] accepted at the HICSS conference (an IT conference ranked "A" at CORE2017).

# **10. Dissemination**

## **10.1. Promoting Scientific Activities**

#### 10.1.1. Scientific Events Organisation

#### 10.1.1.1. Member of the Organizing Committees

Organization co-chair (T. Guyet) of the first conference on "usage of the medico-administrative database of french insured for health research" (http://www.rennes-donnees-sante-2017.fr/)

Organization chair (T. Guyet) of GAST workshop at EGC 2018 and at EGC 2017 (https://gt-gast.irisa.fr/)

Organization co-chair (T. Guyet) of the technical track "Knowledge Extraction from Geographical Data" of the 33rd ACM/SIGAPP Symposium On Applied Computing

Webmaster (L. Galárraga) of the ICDE 2018 conference (https://icde2018.org/)

#### 10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

AAAI Conference on Artificial Intelligence 2017, 2018 (T. Guyet)

Conférence Nationale sur les Applications Pratiques de l'Intelligence Artificielle (APIA) 2017 (C. Largouët)

IEEE International Conference on Data Science and Advanced Analytics (DSAA) 2017 (A. Termier) European Conference Dedicated to the Future Use of ICT in the Agri-food Sector, Bioresource and Biomass sector (EFITA) 2017 (A. Termier, C. Largouët)

Extraction et Gestion de Connaissances (EGC) 2018 (T. Guyet, R. Quiniou, A. Termier).

European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (EMLCPKDD) 2017 (E. Fromont)

International Conference on Data Mining (ICDM) 2017 (A. Termier)

Conference on Intelligent Data Analysis (IDA) 2017 (E. Fromont)

International Joint Conference on Artificial Intelligence (IJCAI) 2017 (T. Guyet, A. Termier)

Symposium on Information Management and Big Data (SimBig) 2017 (T. Guyet)

10.1.2.2. Reviewer

Y. Dauxais, C. Gautrais: IJCAI 2017, ICDM 2017, EGC 2018

E. Fromont: ECMLPKDD 2017, AAAI 2017, IDA2017, ECMLPKDD demo track 2017

L. Galárraga: The Web Conference 2018, WebDB workshop 2017, AKBC workshop 2017

T. Guyet: IJCAI 2017, ICDM 2017, AAAI 2018, AAAI 2017

- R. Quiniou: IJCAI 2017, ICDM 2017, AAAI 2018
- A. Termier: EuroPar 2017

#### 10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

E. Fromont: Machine Learning and Data Mining ECMLPKDD special issue

T. Guyet: Revue d'Intelligence Artificielle (RIA)

10.1.3.2. Reviewer - Reviewing Activities

E. Fromont: Data Mining and Knowledge Discovery, Machine Learning Journal

L. Galárraga: Data Mining and Knowledge Discovery, Artificial Intelligence Review, Semantic Web Journal

C. Gautrais: Data Mining and Knowledge Discovery

T. Guyet: Journal of Biomedical Informatics, Journal of Intelligent Information Systems, Artificial Intelligence Review

C. Largouët: Natural Computing, International Journal of Agricultural and Environmental Information Systems (IJAEIS)

A. Termier: Knowledge and Information Systems, The Very Large DataBases Journal

#### 10.1.4. Invited Talks

E. Fromont gave an invited talk for the Machine Learning seminar of the KULEUVEN University in May 2017, Ostende, Belgium.

A. Termier gave an invited talk at KU Leuven on 09/06/2017.

E. Fromont gave an invited talk "Big Data and Business" summer school oragnised by the IRIXYS consortium and ATOS, Chiemsee, Germany.

A. Termier gave an invited talk at the esaconnect event of ESA agriculture school (Angers), on 26/10/2017.

L. Galárraga gave an invited talk at the LTCI Data Science Seminar (https://ltci.telecom-paristech. fr/data-science-seminar/) at Télécom ParisTech on 16/11/2017.

T. Guyet gave an invited talk at the SPECIF Campus association (https://www.specifcampus.fr/les-supports-de-la-journee-du-24-novembre-2016/) on 24/11/2017.

#### 10.1.5. Leadership within the Scientific Community

A. Termier is the representative of Inria for the #DigitAg Convergence Institute

#### 10.1.6. Scientific Expertise

Evaluation of one project proposal for IFREMER: T. Guyet

Evaluation of two projects proposals for ANR: A. Termier

Comité d'évaluation scientifique ANR CE23 - Données, Connaissances, Big data, Contenus multimédias, Intelligence Artificielle: E. Fromont

Groupe de travail national "infrastructure de recherche pour l'Intelligence Artificielle" (demandé par Allistene): E. Fromont

#### 10.1.7. Research Administration

Member of INRA CEI (Engineers Evaluation Committee): T. Guyet Member of the scientific board of department EA of INRA: A. Termier Member of the scientific board of Agrocampus Ouest - COREGE: C. Largouët

# **10.2. Teaching - Supervision - Juries**

### 10.2.1. Teaching

Many members of the project-team LACODAM are also faculty members and are actively involved in computer science teaching programs in ISTIC, INSA and Agrocampus-Ouest. Besides these usual teaching activities, LACODAM is involved in the following programs:

A. Termier, DMV Module: Data Mining and Visualization, 18h, Master 2, Istic, Univ. Rennes 1

L. Bonneau and C. Largouët, DataViz with R, 10h, Master Datascience, Agrocampus Ouest Rennes

L. Bonneau and C. Largouët, Computer Science for BigData, 30h, Master Datascience, Agrocampus Ouest Rennes

C. Largouët, Scientific Programming, Master 1, Agrocampus Ouest Rennes

C. Largouët, Data Management, Master 1, Agrocampus Ouest Rennes

#### 10.2.2. Supervision

PhD in progress: Maël Guillemé, "New data mining approaches for improving energy consumption in factory", 03/10/2016, A. Termier, V. Masson, L. Rozé and R. Quiniou

PhD in progress: Clément Gautrais, "Mining massive data from client purchases", 01/10/2015, A. Termier, P. Cellier, T. Guyet and R. Quiniou

PhD in progress: Yann Dauxais, "Query-language for care-pathway mining and analysis", 01/02/2015, D. Gross-Amblard, T. Guyet, A. Happe

PhD in progress: Colin Leverger, "Cluster resources optimization through forecasting and management of metric time series", 01/10/2017, T. Guyet, S. Malinowski, R. Marguerie, A. Termier

PhD in progress: Alban Siffer, "DataMining approaches for cyber attack detection", 05/2016, P.-A. Fouque, A. Termier, C. Largouët

PhD in progress: Kevin Fauvel, "Using data mining techniques for improving dairy management", 10/2017, V. Masson, P. Faverdin, A. Termier

PhD in progress: Raphaël Gauthier, "Modelling of nutrient utilization and precision feeding of lactating sows", 11/2017, C. Largouët, J.-Y. Dourmad

## 10.2.3. Juries

Committee member of Georges Nassopoulos' PhD defense on 22/05/2017 (Université de Nantes): R. Quiniou

Committee member of Chemseddine NABTI's PhD defense (Université Claude Bernard Lyon 1): A. Termier

Reviewer of Vladimir Dzyuba's Phd (KU Leuven): A. Termier

Committee member of Hoang Son Pham's PhD defense (Université de Rennes 1): A. Termier

Thesis advisory committee member of Jean Coquet (Univ. Rennes 1): A. Termier

Thesis advisory committee member of Mathilde Chen (INRA): A. Termier

Thesis advisory committee member of Benoit Bellot (INRA/IGEPP): T. Guyet

Thesis advisory committee member of Zhi Cheng (UNC/PPME): T. Guyet

Thesis advisory committee member of Vanel Siyou (Univ. Blaise Pascal/LIMOS): T. Guyet

Committee member of Asma Dachraoui's PhD defense on 31/01/2017 (ABIES/AgroParisTech): T. Guyet

Member of the associate professor hiring commitee on 12/05/2017 (LINK/AgroParisTech) : T. Guyet

Committee member of Jordy Salmon-Monviola's PhD defense on 05/04/2017 (AgroCampus Ouest): V. Masson

Committee member (as co-supervisor) of Damien Fourure on 12/12/2017 (Univ. Saint-Etienne) : E Fromont

Committee member of Van-Tinh TRAN, (University of Lyon) on 11/07/2017 : E. Fromont

Committee member of Maxime Gasse, (University of Lyon on 13/01/2017: E. Fromont

Committee member of Pauline Wauquier, (University of Lille 3) on 29/05/2017 : E. Fromont

Committee member of Ouadie Gharroudi, (University of Lyon) on 21/12/2017 : E. Fromont

Thesis advisory committee member of Mohamed Ali Hammal on 30/04/2017 (University of Lyon), Maxime CHABERT on 11/04/2017 and 2/06/2017 (University of Lyon), Tuan Nguyen on 21/06/2017 (Université Savoie Mont Blanc), Jean-Jacques Ponciano on 17/05/2017 (Univ. Saint-Etienne), Julien Tissier, Jordan Fréry, Carlos Arango, Dennis Diefenbach on 12/07/2017 (Univ. Saint-Etienne): E. Fromont

## **10.3.** Popularization

M.-O. Cordier is editorial board member of Interstices webzine.

C. Largouët and L. Galárraga participated in a science diffusion talk at the educational institution "Assomption" on 20/10/2017.

L. Galárraga participated as AI specialist in a discussion forum at the general election assembly of the "Pôle Images et Reseaux" on 23/11/2017.

T. Guyet gave a public talk about Artificial Intelligence and Data Mining in MDA Rennes (FranceIA events) on 01/03/2017.

A. Termier gave a public talk about Artificial Intelligence at the Esaconnect event on 26/10/2017 https://www.chaire-mutations-agricoles.com/evenements/esaconnect-2017/programme/panorama-de-lintelligence-artificielle-en-2017-alexandre-termier/link.

A. Termier was interviewed for the December podcast of the Interstices webzine.

# 11. Bibliography

## Major publications by the team in recent years

[1] M. GEBSER, T. GUYET, R. QUINIOU, J. ROMERO, T. SCHAUB. Knowledge-based Sequence Mining with ASP, in "IJCAI 2016- 25th International joint conference on artificial intelligence", New-york, United States, AAAI, July 2016, 8, https://hal.inria.fr/hal-01327363. [2] A. SIFFER, P.-A. FOUQUE, A. TERMIER, C. LARGOUËT. Anomaly Detection in Streams with Extreme Value Theory, in "KDD 2017 - Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining", Halifax, Canada, August 2017 [DOI: 10.1145/3097983.3098144], https:// hal.archives-ouvertes.fr/hal-01640325.

## **Publications of the year**

#### **Articles in International Peer-Reviewed Journal**

- [3] T. BOUADI, M.-O. CORDIER, P. MOREAU, R. QUINIOU, J. SALMON-MONVIOLA, C. GASCUEL-ODOUX.A data warehouse to explore multidimensional simulated data from a spatially distributed agro-hydrological model to improve catchment nitrogen management, in "Environmental Modelling and Software", November 2017, vol. 97, p. 229 - 242 [DOI: 10.1016/J.ENVSOFT.2017.07.019], https://hal.inria.fr/hal-01597840.
- [4] E. DREZEN, T. GUYET, A. HAPPE.From medico-administrative databases analysis to care trajectories analytics: an example with the French SNDS, in "Fundamental and Clinical Pharmacology", September 2017 [DOI: 10.1111/FCP.12323], https://hal.inria.fr/hal-01631802.
- [5] C. LARGOUËT, O. KRICHEN, Y. ZHAO. Extended Automata for Temporal Planning of Interacting Agents, in "International Journal of Monitoring and Surveillance Technologies Research", April 2017, vol. 5, n<sup>o</sup> 1, p. 30 - 48 [DOI: 10.4018/IJMSTR.2017010102], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01640137.
- [6] V. LEROY, M. KIRCHGESSNER, A. TERMIER, S. AMER-YAHIA. TopPI: An efficient algorithm for item-centric mining, in "Information Systems", 2017, vol. 64, p. 104 - 118 [DOI: 10.1016/J.IS.2016.09.001], https:// hal.archives-ouvertes.fr/hal-01479067.

#### **International Conferences with Proceedings**

- [7] Y. DAUXAIS, D. GROSS-AMBLARD, T. GUYET, A. HAPPE. Extraction de chroniques discriminantes, in "Extraction et Gestion des Connaissances (EGC)", Grenoble, France, January 2017, https://hal.inria.fr/hal-01413473.
- [8] Y. DAUXAIS, T. GUYET, D. GROSS-AMBLARD, A. HAPPE.Discriminant chronicles mining: Application to care pathways analytics, in "Artificial Intelligence in Medicine", Vienna, Austria, 16th Conference on Artificial Intelligence in Medicine, June 2017, https://arxiv.org/abs/1709.03309 [DOI: 10.1007/978-3-319-59758-46], https://hal.archives-ouvertes.fr/hal-01568929.
- [9] C. GAUTRAIS, P. CELLIER, R. QUINIOU, A. TERMIER. *Topic Signatures in Political Campaign Speeches*, in "EMNLP 2017 - Conference on Empirical Methods in Natural Language Processing", Copenhagen, Denmark, Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, September 2017, https://hal.archives-ouvertes.fr/hal-01640498.
- [10] C. GAUTRAIS, R. QUINIOU, P. CELLIER, T. GUYET, A. TERMIER. Purchase Signatures of Retail Customers, in "PAKDD 2017 - The Pacific-Asia Conference on Knowledge Discovery and Data Mining", Jeju, South Korea, Pacific-Asia Conference on Knowledge Discovery and Data Mining, May 2017, https://hal.archivesouvertes.fr/hal-01639795.
- [11] M. GUILLEME, L. ROZÉ, V. MASSON, C. CARTON, R. QUINIOU, A. TERMIER. Improving time-series rule matching performance for detecting energy consumption patterns, in "DARE 2017 5th International

Workshop on Data Analytics for Renewable Energy Integration", Skopje, Macedonia, Springer, September 2017, vol. 10691, p. 59-71 [*DOI* : 10.1007/978-3-319-71643-5\_6], https://hal.inria.fr/hal-01654890.

- [12] T. GUYET, A. HAPPE, Y. DAUXAIS. Declarative Sequential Pattern Mining of Care Pathways, in "Conference on Artificial Intelligence in Medicine in Europe", Vienna, Austria, 16th Conference on Artificial Intelligence in Medicine, June 2017, vol. 24, p. 1161 - 266, https://arxiv.org/abs/1707.08342 [DOI: 10.1007/978-3-319-59758-4\_29], https://hal.inria.fr/hal-01569023.
- [13] T. GUYET, R. QUINIOU, V. MASSON. *Mining relevant interval rules*, in "International Conference on Formal Concept Analysis", Rennes, France, Supplementary proceedings of International Conference on Formal Concept Analysis (ICFCA), June 2017, https://arxiv.org/abs/1709.03267, https://hal.inria.fr/hal-01584981.
- [14] S. KELLY. A Communication Model that Bridges Knowledge Delivery between Data Miners and Domain Users , in "51th Hawaii International Conference on System Sciences (HICSS)", Hawaii, United States, January 2018, https://hal.archives-ouvertes.fr/hal-01651737.
- [15] A. SAMET, T. GUYET, B. NEGREVERGNE, T.-T. DAO, T. NHA HOANG, M.-C. HO BA THO.*Expert Opinion Extraction from a Biomedical Database*, in "Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)", Lugano, Switzerland, Proceedings of 14th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Springer, July 2017, vol. 31, n<sup>o</sup> LNCS 10369, p. 1 12, https://arxiv.org/abs/1709.03270 [DOI : 10.1016/S0888-613X(02)00066-X], https://hal.inria.fr/hal-01584984.
- [16] A. SIFFER, P.-A. FOUQUE, A. TERMIER, C. LARGOUËT. Anomaly Detection in Streams with Extreme Value Theory, in "KDD 2017 - Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining", Halifax, Canada, August 2017 [DOI : 10.1145/3097983.3098144], https:// hal.archives-ouvertes.fr/hal-01640325.
- [17] K. TSESMELI, M. BOUMGHAR, T. GUYET, R. QUINIOU, L. PIERRE. Fouille de motifs temporels négatifs, in "EGC 2018 - 18ème Conférence Internationale sur l'Extraction et la Gestion des Connaissances", Paris, France, January 2018, p. 1-6, https://hal.inria.fr/hal-01657540.

#### **Conferences without Proceedings**

- [18] P. BESNARD, T. GUYET, V. MASSON. Admissible generalizations of examples as rules, in "11e Journées d'Intelligence Artificielle Fondamentale", Caen, France, July 2017, https://hal.archives-ouvertes.fr/hal-01576047.
- [19] C. GAUTRAIS, Y. DAUXAIS, M. GUILLEME. Multi-Plant Photovoltaic Energy Forecasting Challenge: Second place solution, in "Discovery Challenges co-located with European Conference on Machine Learning -Principle and Practice of Knowledge Discovery in Database", Skopje, Macedonia, September 2017, https:// hal.archives-ouvertes.fr/hal-01639813.
- [20] A. SAMET, T. GUYET, B. NEGREVERGNE. *Mining rare sequential patterns with ASP*, in "ILP 2017 27th International Conference on Inductive Logic Programming", Orléans, France, September 2017, https://hal. archives-ouvertes.fr/hal-01569582.

#### Scientific Books (or Scientific Book chapters)

- [21] M.-O. CORDIER, P. DAGUE, Y. PENCOLÉ, L. TRAVÉ-MASSUYÈS. *Diagnosis and supervision: model-based approaches*, in "A guided tour of artificial intelligence research", H. P. PIERRE MARQUIS (editor), Knowledge representation and reasoning, Springer, 2018, n<sup>O</sup> 1, https://hal.archives-ouvertes.fr/hal-01483436.
- [22] T. GUYET, Y. MOINARD, R. QUINIOU, T. SCHAUB.Efficiency Analysis of ASP Encodings for Sequential Pattern Mining Tasks, in "Advances in Knowledge Discovery and Management", B. PINAUD, F. GUILLET, B. CREMILLEUX, C. DE RUNZ (editors), Springer, October 2017, vol. 7, p. 41–81, https://arxiv.org/abs/1711. 05090, https://hal.inria.fr/hal-01631879.
- [23] A. NAPOLI, A. TERMIER. La fouille de données, in "Les Big Data à découvert", M. BOUZEGHOUB, R. MOSSERI (editors), CNRS Editions, 2017, p. 1-3, https://hal.inria.fr/hal-01673437.

#### **Scientific Popularization**

- [24] M.-O. CORDIER.*Regard sur « Le mythe de la Singularité. Faut-il craindre l'intelligence artificielle ? »*, in "Interstices", June 2017, https://hal.inria.fr/hal-01616345.
- [25] A. TERMIER, J. JONGWANE. Vers une démocratisation des outils pour l'exploration de données ?, in "Interstices", December 2017, https://hal.inria.fr/hal-01688785.

#### **References in notes**

- [26] L. BREIMAN. Bagging predictors, in "Machine Learning", Aug 1996, vol. 24, n<sup>o</sup> 2, p. 123–140, https://doi. org/10.1007/BF00058655.
- [27] S. COLAS, C. COLLIN, P. PIRIOU, M. ZUREIK. Association between total hip replacement characteristics and 3-year prosthetic survivorship: A population-based study, in "JAMA Surgery", 2015, vol. 150, n<sup>o</sup> 10, p. 979–988.
- [28] G. C. GARRIGA, P. KRALJ, N. LAVRAC. Closed Sets for Labeled Data, in "Journal of Machine Learning Research", 2008, vol. 9, p. 559–580, http://doi.acm.org/10.1145/1390681.1390700.
- [29] M. KAYTOUE, S. O. KUZNETSOV, A. NAPOLI. Revisiting numerical pattern mining with formal concept analysis, in "IJCAI Proceedings-International Joint Conference on Artificial Intelligence", 2011, vol. 22, n<sup>o</sup> 1, 1342.
- [30] N. LANDWEHR, M. HALL, E. FRANK. Logistic model trees, in "Machine learning", 2005, vol. 59, n<sup>o</sup> 1-2, p. 161–205.
- [31] G. MOULIS, M. LAPEYRE-MESTRE, A. PALMARO, G. PUGNET, J.-L. MONTASTRUC, L. SAILLER.*French health insurance databases: What interest for medical research?*, in "La Revue de Médecine Interne", 2015, vol. 36, n<sup>0</sup> 6, p. 411 417.
- [32] E. NOWAK, A. HAPPE, J. BOUGET, F. PAILLARD, C. VIGNEAU, P.-Y. SCARABIN, E. OGER. Safety of Fixed Dose of Antihypertensive Drug Combinations Compared to (Single Pill) Free-Combinations: A Nested Matched Case–Control Analysis, in "Medicine", 2015, vol. 94, n<sup>o</sup> 49, e2229.

- [33] E. POLARD, E. NOWAK, A. HAPPE, A. BIRABEN, E. OGER.Brand name to generic substitution of antiepileptic drugs does not lead to seizure-related hospitalization: a population-based case-crossover study, in "Pharmacoepidemiology and drug safety", 2015, vol. 24, n<sup>o</sup> 11, p. 1161–1169.
- [34] R. E. SCHAPIRE. *The Boosting Approach to Machine Learning: An Overview*, Springer New York, New York, NY, 2003, p. 149–171.
- [35] M. SHOKOOHI-YEKTA, Y. CHEN, B. CAMPANA, B. HU, J. ZAKARIA, E. KEOGH. Discovery of meaningful rules in time series, in "Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining", ACM, 2015, p. 1085–1094.
- [36] R. WANG, Y.-L. HE, C.-Y. CHOW, F.-F. OU, J. ZHANG.Learning ELM-tree from big data based on uncertainty reduction, in "Fuzzy Sets and Systems", 2015, vol. 258, p. 79–100.

# **Project-Team LAGADIC**

# Visual servoing in robotics, computer vision, and augmented reality

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: CNRS Institut national des sciences appliquées de Rennes Université Rennes 1

RESEARCH CENTERS Rennes - Bretagne-Atlantique Sophia Antipolis - Méditerranée

THEME Robotics and Smart environments

# **Table of contents**

1.	Personnel	••••	569
2.			
3.	Research	Program	5/1
	3.1. V1S	ual Servoing	571
	3.2. V1S	ual Tracking	573
	3.3. $SLA$	AM na Madalina and Undarstandina	575
4	3.4. Sce	ne Modeling and Understanding	5/5
4. -	Applicatio		5/4
э. С	Hignlight	s of the year	
0.	New Solly	vare and Plauorms	3/3
	0.1. DID.	2ntml Slom	575
	0.2. DE		575
	0.3. Hai		575
	0.4. Per		575
	0.3. SIN		576
	$\begin{array}{ccc} 0.0. & USI \\ 6.7 & ViSI \\ \end{array}$		576
	0.7. VIS	r forms	570
	0.0. Fla	Pahat Vision Distform	577
	682	Mobile Pohots	577
	6.8	2.1 Indeer Mahile Pabets	577
	6.8	2.1. Indoor Webicles	570
	683	Medical Robotic Platform	579
	684	Humanoid Robots	579
	685	Unmanned Aerial Vehicles (UAVs)	582
7	New Resu	lts	582
	7 1 Vis	ual Percention	582
	7.1.1.	Visual Tracking for Motion Capture	582
	7.1.2	Object 3D Tracking based on Depth Information and CAD Model	584
	7.1.3.	General Model-based Tracker	584
	7.1.4.	3D Localization for Airplane Landing	584
	7.1.5.	Extrinsic Calibration of Multiple RGB-D Cameras	584
	7.1.6.	Scene Registration with Large Convergence Domain	585
	7.1.7.	Scene Semantization based on Deep Learning Approach	585
	7.1.8.	Online Localization and Mapping for UAVs	586
	7.1.9.	Reflectance and Illumination Estimation for Realistic Augmented Reality	586
	7.1.10.	Optimal Active Sensing Control	586
	7.2. Sensor-based Robot Control		587
	7.2.1.	Determining Singularity Configurations in IBVS	587
	7.2.2.	Visual Servoing through Mirror Reflection	587
	7.2.3.	Visual Servoing of Humanoid Robots	587
	7.2.4.	Model Predictive Visual Servoing	587
	7.2.5.	Model Predictive Control for Visual Servoing of a UAV	588
	7.2.6.	UAVs in Physical Interaction with the Environment	588
	7.2.7.	Visual Servoing for Steering Simulation Agents	588
	7.2.8.	Direct Visual Servoing	589
	7.3. Me	dical Robotics	589
	7.3.1.	Visual Servoing using Wavelet and Shearlet Transforms	589
	7.3.2.	3D Steering of Flexible Needle by Ultrasound Visual Servoing	589

	7.3.3. Robotic Assistance for Ultrasound Elastography by Visual Servoing, F	Force Control and
	Teleoperation	590
	7.3.4. Haptic Guidance of a Biopsy Needle	590
	7.4. Teleoperation	590
	7.4.1. Shared Control for Remote Manipulation	590
	7.4.2. Wearable haptics	591
	7.5. Navigation of Mobile Robots	593
	7.5.1. Visual Navigation from an Image Memory	593
	7.5.2. Robot-Human Interactions during Locomotion	593
	7.5.3. Semi-Autonomous Control of a Wheelchair for Navigation Assistance	593
	7.5.4. Wheelchair Kinematics and Dynamics Modeling for Shared Control	594
	7.5.5. Wheelchair Autonomous Navigation for Fall Prevention	594
	7.5.6. Robotic Platform for Assistance to People with Reduce Mobility	594
	7.6. Multi-robot and Crowd Motion Control	595
	7.6.1. Rigidity-based Methods for Formation Control	595
	7.6.2. Cooperative Localization using Interval Analysis	595
8.	Bilateral Contracts and Grants with Industry	<b>595</b>
	8.1. Bilateral Contracts with Industry	595
	8.1.1. Robocortex	595
	8.1.2. ABB	596
	8.1.3. IRT b<>com	596
	8.2. Bilateral Grants with Industry	596
	8.2.1. Technicolor	596
	8.2.2. Realyz	596
	8.2.3. Pôle Saint Hélier	596
	8.2.4. Axyn	596
9.	Partnerships and Cooperations	<b>597</b>
	9.1. Regional Initiatives	597
	9.1. Regional Initiatives 9.1.1. ARED DeSweep	597 597
	<ul><li>9.1. Regional Initiatives</li><li>9.1.1. ARED DeSweep</li><li>9.1.2. ARED Locoflot</li></ul>	597 597 597
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> </ul>	597 597 597 597 597
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> </ul>	597 597 597 597 597 597
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> </ul>	597 597 597 597 597 597 597
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> <li>9.1.6. IRT Jules Verne Mascot</li> </ul>	597 597 597 597 597 597 597 597
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> <li>9.1.6. IRT Jules Verne Mascot</li> <li>9.1.7. IRT b&lt;&gt;com NeedleWare</li> </ul>	597 597 597 597 597 597 597 597 597
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> <li>9.1.6. IRT Jules Verne Mascot</li> <li>9.1.7. IRT b&lt;&gt;com NeedleWare</li> <li>9.1.8. Prisme</li> </ul>	597 597 597 597 597 597 597 597 597 598
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> <li>9.1.6. IRT Jules Verne Mascot</li> <li>9.1.7. IRT b&lt;&gt;com NeedleWare</li> <li>9.1.8. Prisme</li> <li>9.2. National Initiatives</li> </ul>	597 597 597 597 597 597 597 597 598 598
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> <li>9.1.6. IRT Jules Verne Mascot</li> <li>9.1.7. IRT b&lt;&gt;com NeedleWare</li> <li>9.1.8. Prisme</li> <li>9.2. National Initiatives</li> <li>9.2.1. France Life Imaging WP3-FLI ANFEET</li> </ul>	597 597 597 597 597 597 597 597 598 598 598
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> <li>9.1.6. IRT Jules Verne Mascot</li> <li>9.1.7. IRT b&lt;&gt;com NeedleWare</li> <li>9.1.8. Prisme</li> <li>9.2. National Initiatives</li> <li>9.2.1. France Life Imaging WP3-FLI ANFEET</li> <li>9.2.2. ANR Contint Visioland</li> </ul>	597 597 597 597 597 597 597 597 598 598 598 598
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> <li>9.1.6. IRT Jules Verne Mascot</li> <li>9.1.7. IRT b&lt;&gt;com NeedleWare</li> <li>9.1.8. Prisme</li> <li>9.2. National Initiatives</li> <li>9.2.1. France Life Imaging WP3-FLI ANFEET</li> <li>9.2.2. ANR Contint Visioland</li> <li>9.2.3. ANR Contint Entracte</li> </ul>	597 597 597 597 597 597 597 597 598 598 598 598 598
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> <li>9.1.6. IRT Jules Verne Mascot</li> <li>9.1.7. IRT b&lt;&gt;com NeedleWare</li> <li>9.1.8. Prisme</li> <li>9.2. National Initiatives</li> <li>9.2.1. France Life Imaging WP3-FLI ANFEET</li> <li>9.2.2. ANR Contint Visioland</li> <li>9.2.3. ANR Contint Entracte</li> <li>9.2.4. ANR JCJC Percolation</li> </ul>	597 597 597 597 597 597 597 597 598 598 598 598 598 598
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> <li>9.1.6. IRT Jules Verne Mascot</li> <li>9.1.7. IRT b&lt;&gt;com NeedleWare</li> <li>9.1.8. Prisme</li> <li>9.2. National Initiatives</li> <li>9.2.1. France Life Imaging WP3-FLI ANFEET</li> <li>9.2.2. ANR Contint Visioland</li> <li>9.2.3. ANR Contint Entracte</li> <li>9.2.4. ANR JCJC Percolation</li> <li>9.2.5. ANR JCJC SenseFly</li> </ul>	597 597 597 597 597 597 597 597 598 598 598 598 598 598
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> <li>9.1.6. IRT Jules Verne Mascot</li> <li>9.1.7. IRT b&lt;&gt;com NeedleWare</li> <li>9.1.8. Prisme</li> <li>9.2. National Initiatives</li> <li>9.2.1. France Life Imaging WP3-FLI ANFEET</li> <li>9.2.2. ANR Contint Visioland</li> <li>9.2.3. ANR Contint Entracte</li> <li>9.2.4. ANR JCJC Percolation</li> <li>9.2.5. ANR JCJC SenseFly</li> <li>9.2.6. ANR PLaTINUM</li> </ul>	597 597 597 597 597 597 597 597 598 598 598 598 598 598 598
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> <li>9.1.6. IRT Jules Verne Mascot</li> <li>9.1.7. IRT b&lt;&gt;com NeedleWare</li> <li>9.1.8. Prisme</li> <li>9.2. National Initiatives</li> <li>9.2.1. France Life Imaging WP3-FLI ANFEET</li> <li>9.2.2. ANR Contint Visioland</li> <li>9.2.3. ANR Contint Entracte</li> <li>9.2.4. ANR JCJC Percolation</li> <li>9.2.5. ANR JCJC SenseFly</li> <li>9.2.6. ANR PLaTINUM</li> <li>9.2.7. BPI Romeo 2</li> </ul>	597 597 597 597 597 597 597 597 598 598 598 598 598 598 598 598 598
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> <li>9.1.6. IRT Jules Verne Mascot</li> <li>9.1.7. IRT b&lt;&gt;com NeedleWare</li> <li>9.1.8. Prisme</li> <li>9.2.1. France Life Imaging WP3-FLI ANFEET</li> <li>9.2.2. ANR Contint Visioland</li> <li>9.2.3. ANR Contint Entracte</li> <li>9.2.4. ANR JCJC Percolation</li> <li>9.2.5. ANR JCJC SenseFly</li> <li>9.2.6. ANR PLaTINUM</li> <li>9.2.7. BPI Romeo 2</li> <li>9.2.8. Equipex Robotex</li> </ul>	597 597 597 597 597 597 597 597 598 598 598 598 598 598 598 598 598 598
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> <li>9.1.6. IRT Jules Verne Mascot</li> <li>9.1.7. IRT b&lt;&gt;com NeedleWare</li> <li>9.1.8. Prisme</li> <li>9.2. National Initiatives</li> <li>9.2.1. France Life Imaging WP3-FLI ANFEET</li> <li>9.2.2. ANR Contint Visioland</li> <li>9.2.3. ANR Contint Entracte</li> <li>9.2.4. ANR JCJC Percolation</li> <li>9.2.5. ANR JCJC SenseFly</li> <li>9.2.6. ANR PLaTINUM</li> <li>9.2.7. BPI Romeo 2</li> <li>9.2.8. Equipex Robotex</li> <li>9.3. European Initiatives</li> </ul>	597 597 597 597 597 597 597 597 598 598 598 598 598 598 598 598 598 598
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> <li>9.1.6. IRT Jules Verne Mascot</li> <li>9.1.7. IRT b&lt;&gt;com NeedleWare</li> <li>9.1.8. Prisme</li> <li>9.2. National Initiatives</li> <li>9.2.1. France Life Imaging WP3-FLI ANFEET</li> <li>9.2.2. ANR Contint Visioland</li> <li>9.2.3. ANR Contint Entracte</li> <li>9.2.4. ANR JCJC Percolation</li> <li>9.2.5. ANR JCJC SenseFly</li> <li>9.2.6. ANR PLaTINUM</li> <li>9.2.7. BPI Romeo 2</li> <li>9.2.8. Equipex Robotex</li> <li>9.3.1. FP7 &amp; H2020 Projects</li> </ul>	597 597 597 597 597 597 597 597 598 598 598 598 598 598 598 598 598 598
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> <li>9.1.6. IRT Jules Verne Mascot</li> <li>9.1.7. IRT b&lt;&gt;com NeedleWare</li> <li>9.1.8. Prisme</li> <li>9.2. National Initiatives</li> <li>9.2.1. France Life Imaging WP3-FLI ANFEET</li> <li>9.2.2. ANR Contint Visioland</li> <li>9.2.3. ANR Contint Entracte</li> <li>9.2.4. ANR JCJC Percolation</li> <li>9.2.5. ANR JCJC SenseFly</li> <li>9.2.6. ANR PLaTINUM</li> <li>9.2.7. BPI Romeo 2</li> <li>9.2.8. Equipex Robotex</li> <li>9.3.1. FP7 &amp; H2020 Projects</li> <li>9.3.1.1. FP7 Space RemoveDEBRIS</li> </ul>	597 597 597 597 597 597 597 597 598 598 598 598 598 598 598 598 598 598
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> <li>9.1.6. IRT Jules Verne Mascot</li> <li>9.1.7. IRT b&lt;&gt;com NeedleWare</li> <li>9.1.8. Prisme</li> <li>9.2. National Initiatives</li> <li>9.2.1. France Life Imaging WP3-FLI ANFEET</li> <li>9.2.2. ANR Contint Visioland</li> <li>9.2.3. ANR Contint Entracte</li> <li>9.2.4. ANR JCJC Percolation</li> <li>9.2.5. ANR JCJC SenseFly</li> <li>9.2.6. ANR PLaTINUM</li> <li>9.2.7. BPI Romeo 2</li> <li>9.2.8. Equipex Robotex</li> <li>9.3.1. FP7 &amp; H2020 Projects</li> <li>9.3.1.1. FP7 Space RemoveDEBRIS</li> <li>9.3.1.2. H2020 ICT Comanoid</li> </ul>	597 597 597 597 597 597 597 597 598 598 598 598 598 598 598 598 598 598
	<ul> <li>9.1. Regional Initiatives</li> <li>9.1.1. ARED DeSweep</li> <li>9.1.2. ARED Locoflot</li> <li>9.1.3. ARED Mod4Nav</li> <li>9.1.4. "Equipement mi-lourd Rennes Métropole"</li> <li>9.1.5. "Allocation d'installation scientifique"</li> <li>9.1.6. IRT Jules Verne Mascot</li> <li>9.1.7. IRT b&lt;&gt;com NeedleWare</li> <li>9.1.8. Prisme</li> <li>9.2. National Initiatives</li> <li>9.2.1. France Life Imaging WP3-FLI ANFEET</li> <li>9.2.2. ANR Contint Visioland</li> <li>9.2.3. ANR Contint Entracte</li> <li>9.2.4. ANR JCJC Percolation</li> <li>9.2.5. ANR JCJC SenseFly</li> <li>9.2.6. ANR PLaTINUM</li> <li>9.2.7. BPI Romeo 2</li> <li>9.2.8. Equipex Robotex</li> <li>9.3.1. FP7 &amp; H2020 Projects</li> <li>9.3.1.1. FP7 Space RemoveDEBRIS</li> <li>9.3.1.2. H2020 ICT Comanoid</li> <li>9.3.1.3. H2020 ICT Romans</li> </ul>	597 597 597 597 597 597 597 597 598 598 598 598 598 598 598 598 598 598

	9.3.3. Collaborations with European Partners	601
	9.3.3.1. ANR Opmops	601
	9.3.3.2. iProcess	602
	9.4. International Initiatives	602
	9.4.1. Inria Associate Teams Not Involved in an Inria International Labs	602
	9.4.1.1. SIMS	602
	9.4.1.2. ISI4NAVE	602
	9.4.2. Participation in International Programs	603
	9.5. International Research Visitors	603
	9.5.1. Visits of International Scientists	603
	9.5.2. Visits to International Teams	603
10.	Dissemination	604
	10.1. Promoting Scientific Activities	604
	10.1.1. Scientific Events Organisation	604
	10.1.1.1. General Chair, Scientific Chair	604
	10.1.1.2. Member of the Organizing Committees	604
	10.1.1.3. Chair of Conference Program Committees	604
	10.1.1.4. Member of the Conference Program Committees	604
	10.1.1.5. Reviewer	604
	10.1.2. Journal	605
	10.1.2.1. Member of the Editorial Boards	605
	10.1.2.2. Reviewer - Reviewing Activities	605
	10.1.3. Invited Talks	605
	10.1.4. Leadership within the Scientific Community	606
	10.1.5. Scientific Expertise	606
	10.1.6. Research Administration	606
	10.2. Teaching - Supervision - Juries	607
	10.2.1. Teaching	607
	10.2.2. Supervision	608
	10.2.3. External Ph.D. and HdR Juries	609
	10.3. Popularization	609
11.	Bibliography	610

# **Project-Team LAGADIC**

*Creation of the Project-Team: 2004 December 06, end of the Project-Team: 2017 December 31* **Keywords:** 

## **Computer Science and Digital Science:**

- A5.4. Computer vision
- A5.4.4. 3D and spatio-temporal reconstruction
- A5.4.5. Object tracking and motion analysis
- A5.4.6. Object localization
- A5.4.7. Visual servoing
- A5.6. Virtual reality, augmented reality
- A5.10. Robotics
- A5.10.2. Perception
- A5.10.4. Robot control
- A5.10.5. Robot interaction (with the environment, humans, other robots)
- A5.10.6. Swarm robotics

#### **Other Research Topics and Application Domains:**

- B2.4.3. Surgery
- B2.5. Handicap and personal assistances
- B5.1. Factory of the future
- B5.6. Robotic systems
- B7.2.1. Smart vehicles
- B8.4. Security and personal assistance

# 1. Personnel

#### **Research Scientists**

François Chaumette [Team leader, Inria, Senior Researcher, HDR] Alexandre Krupa [Inria, Researcher, HDR] Claudio Pacchierotti [CNRS, Researcher] Julien Pettré [Inria, Researcher, HDR] Patrick Rives [Inria, Senior Researcher, HDR] Paolo Robuffo Giordano [CNRS, Senior Researcher, HDR] Paolo Salaris [Inria, Researcher]

### **Faculty Members**

Jason Chevrie [Univ de Rennes 1, Assistant Professor, from Sep 2017] Marie Babel [INSA Rennes, Associate Professor, HDR] Vincent Drevelle [Univ de Rennes 1, Associate Professor] Eric Marchand [Univ de Rennes 1, Professor, HDR] Philippe Martinet [Inria, Professor, from Nov 2017, in secondment from Ecole Centrale de Nantes, HDR]

#### **Technical Staff**

Fabien Spindler [Inria, Senior engineer] Thomas Bellavoir [CNRS, until Mar 2017, granted by ANR SenseFly project] Nicolas Cazy [Inria, granted by Inria Hub] Giovanni Claudio [Inria, until Aug 2017, granted by H2020 ICT Comanoid project] Solenne Fortun [INSA Rennes, granted by Prisme project] Pierre-Marie Kerzerho [Inria, until Oct 2017, granted by ANR VisioLand] Pol Mordel [CNRS, granted by H2020 ICT Romans project] Marc Pouliquen [Inria, granted by ADT UsTk] Souriya Trinh [Inria, granted by H2020 ICT Comanoid project] Aurélien Yol [Inria, until Nov 2017, granted by Inria Hub]

#### **PhD Students**

Firas Abi Farraj [CNRS, granted by H2020 Romans project] Quentin Bateux [Univ de Rennes 1, granted by MESR] Aline Baudry [INSA Rennes, granted by MESR and Brittany Council] Florian Berton [Inria, from Nov 2017, granted by ANR Opmops] Jason Chevrie [Univ de Rennes 1, until Aug 2017, granted by ENS Cachan] Quentin Delamare [Univ de Rennes 1, granted by ENS Rennes] Louise Devigne [Pôle Saint Hélier, Cifre grant] Lesley-Ann Duflot [Inria, granted by Femto-ST and Brittany Council] Hadrien Gurnel [Institut de recherche technologique B-com] Dayana Hassan [Axyn, from Mar 2017, Cifre grant] Salma Jiddi [Technicolor, Cifre grant] Ide Flore Kenmogne Fokam [Inria, granted in part by Brittany Council] Nicolas Le Borgne [INSA Rennes, from Oct 2017, granted by EU Intereg Adapt project] Axel Lopez Gandia [Univ de Rennes 1, granted by MESR] Renato José Martins [CNPq, until Oct 2017] Noël Mériaux [Inria, until Sep 2017, granted by ANR VisioLand] Pedro Alfonso Patlan Rosales [Conacyt] Bryan Penin [Inria, granted in part by DGA] Rahaf Rahal [Univ de Rennes 1, from Oct 2017, granted by MESR] Fabrizio Schiano [Inria, Cordi-S grant] Agniva Sengupta [Inria, Cordi-S grant] Muhammad Usman [CNRS, granted by ANR SenseFly]

#### **Post-Doctoral Fellows**

Marco Aggravi [CNRS, from Feb 2017, granted by H2020 ICT Romans project] Marco Cognetti [CNRS, from Feb 2017, granted by H2020 ICT Romans project] Eduardo Fernandez Moral [Inria, granted by ANR Platinum project] Aly Magassouba [CNRS, until Jun 2017, granted by H2020 ICT Romans project]

#### Visiting Scientists

Chinemelu Ezeh [University College of London (UK), Ph.D. student, from May 2017 until Sep 2017] Amir Masoud Ghalamzan Esfahani [University of Birmingham (UK), Ph.D. student, Feb 2017] Devash Hanoomanjee [University College of London (UK), Ph.D. student, from Jun 2017 until Jul 2017] Ekrem Misimi [Sintef (Norway), Researcher, from May 2017 until Nov 2017] Mario Selvaggio [University of Naples (Italy), Ph.D. student, from Nov 2017] Denis Wolf [University of Sao Paulo (Brazil), Professor, until Jul 2017] Kaixiang Zhang [Zhejiang University (China), Ph.D. student, from Nov 2017]

#### Administrative Assistants

Hélène de La Ruée [Inria] Isabelle Strobant [Inria]

# 2. Overall Objectives

# 2.1. Overall Objectives

Historically, research activities of the Lagadic team are concerned with visual servoing, visual tracking, and active vision. Visual servoing consists in using the information provided by a vision sensor to control the movements of a dynamic system. This research topic is at the intersection of the fields of robotics, automatic control, and computer vision. These fields are the subject of profitable research since many years and are particularly interesting by their very broad scientific and application spectrum. Within this spectrum, we focus on the interaction between visual perception and action. This topic is important because it provides an alternative to the traditional Perception-Decision-Action cycle. It is indeed possible to link the perception and action aspects more closely, by directly integrating the measurements provided by a vision sensor in closed loop control laws. Our objective is thus to design strategies of coupling perception and action from images for applications in robotics, computer vision, virtual reality and augmented reality.

This objective is significant, first of all because of the variety and the great number of potential applications to which our work can lead (see Section 4.1). Secondly, it is also significant to be able to raise the scientific aspects associated with these problems, namely modeling of visual features representing the interaction between action and perception in an optimal way, taking into account of complex environments and the specification of high level tasks. We also work to treat new problems provided by imagery systems such as those resulting from an omnidirectional vision sensor or echographic probes. We are finally interested in revisiting traditional problems in computer vision (3D localization) through the visual servoing approach.

Thanks to the arrival of Patrick Rives and his students in the group in April 2012, which makes Lagadic now localized both in Rennes and Sophia Antipolis, the group now also focuses on building consistent representations of the environment that can be used to trigger and execute the robot actions. In its broadest sense, perception requires detecting, recognizing, and localizing elements of the environment, given the limited sensing and computational resources available on the embedded system. Perception is a fundamental issue for both the implementation of reactive behaviors, as is traditionally studied in the group, and the construction of the representations that are used at the task level. Simultaneous Localization and Mapping (SLAM) is thus now one of our research areas.

Among the sensory modalities, computer vision, range finder and odometry are of particular importance and interest for mobile robots due to their availability and extended range of applicability, while ultrasound images and force measurements are both required for our medical robotics applications. The fusion of complementary information provided by different sensors is thus also a central issue for modeling the environment, robot localization, control, and navigation.

Much of the processing must be performed in real time, with a good degree of robustness so as to accommodate with the large variability of the physical world. Computational efficiency and well-posedness of the methods developed are thus constant preoccupations of the group.

# 3. Research Program

# 3.1. Visual Servoing

Basically, visual servoing techniques consist in using the data provided by one or several cameras in order to control the motions of a dynamic system [1]. Such systems are usually robot arms, or mobile robots, but can also be virtual robots, or even a virtual camera. A large variety of positioning tasks, or mobile target tracking, can be implemented by controlling from one to all the degrees of freedom (DoF) of the system. Whatever the sensor configuration, which can vary from one on-board camera on the robot end-effector to several free-standing cameras, a set of visual features has to be selected at best from the image measurements available, allowing to control the desired DoF. A control law has also to be designed so that these visual features s(t)

reach a desired value  $s^*$ , defining a correct realization of the task. A desired planned trajectory  $s^*(t)$  can also be tracked. The control principle is thus to regulate the error vector  $s(t) - s^*(t)$  to zero. With a vision sensor providing 2D measurements, potential visual features are numerous, since 2D data (coordinates of feature points in the image, moments, ...) as well as 3D data provided by a localization algorithm exploiting the extracted 2D features can be considered. It is also possible to combine 2D and 3D visual features to take the advantages of each approach while avoiding their respective drawbacks.

More precisely, a set s of k visual features can be taken into account in a visual servoing scheme if it can be written:

$$\mathbf{s} = \mathbf{s}(\mathbf{x}(\mathbf{p}(t)), \mathbf{a}) \tag{50}$$

where  $\mathbf{p}(t)$  describes the pose at the instant t between the camera frame and the target frame, x the image measurements, and a a set of parameters encoding a potential additional knowledge, if available (such as for instance a coarse approximation of the camera calibration parameters, or the 3D model of the target in some cases).

The time variation of s can be linked to the relative instantaneous velocity  $\mathbf{v}$  between the camera and the scene:

$$\dot{\mathbf{s}} = \frac{\partial \mathbf{s}}{\partial \mathbf{p}} \, \dot{\mathbf{p}} = \mathbf{L}_{\mathbf{s}} \, \mathbf{v} \tag{51}$$

where  $L_s$  is the interaction matrix related to s. This interaction matrix plays an essential role. Indeed, if we consider for instance an eye-in-hand system and the camera velocity as input of the robot controller, we obtain when the control law is designed to try to obtain an exponential decoupled decrease of the error:

$$\mathbf{v}_{c} = -\lambda \widehat{\mathbf{L}_{\mathbf{s}}}^{+} (\mathbf{s} - \mathbf{s}^{*}) - \widehat{\mathbf{L}_{\mathbf{s}}}^{+} \frac{\partial \widehat{\mathbf{s}}}{\partial t}$$
(52)

where  $\lambda$  is a proportional gain that has to be tuned to minimize the time-to-convergence,  $\widehat{\mathbf{L}_s}^+$  is the pseudoinverse of a model or an approximation of the interaction matrix, and  $\frac{\partial s}{\partial t}$  an estimation of the features velocity due to a possible own object motion.

From the selected visual features and the corresponding interaction matrix, the behavior of the system will have particular properties as for stability, robustness with respect to noise or to calibration errors, robot 3D trajectory, etc. Usually, the interaction matrix is composed of highly non linear terms and does not present any decoupling properties. This is generally the case when s is directly chosen as x. In some cases, it may lead to inadequate robot trajectories or even motions impossible to realize, local minimum, tasks singularities, etc. It is thus extremely important to design adequate visual features for each robot task or application, the ideal case (very difficult to obtain) being when the corresponding interaction matrix is constant, leading to a simple linear control system. To conclude in a few words, visual servoing is basically a non linear control problem. Our Holy Grail quest is to transform it into a linear control problem.

Furthermore, embedding visual servoing in the task function approach allows solving efficiently the redundancy problems that appear when the visual task does not constrain all the DoF of the system. It is then possible to realize simultaneously the visual task and secondary tasks such as visual inspection, or joint limits or singularities avoidance. This formalism can also be used for tasks sequencing purposes in order to deal with high level complex applications.

# **3.2. Visual Tracking**

Elaboration of object tracking algorithms in image sequences is an important issue for researches and applications related to visual servoing and more generally for robot vision. A robust extraction and real time spatio-temporal tracking process of visual cues is indeed one of the keys to success of a visual servoing task. If fiducial markers may still be useful to validate theoretical aspects in modeling and control, natural scenes with non-cooperative objects and subject to various illumination conditions have to be considered for addressing large scale realistic applications.

Most of the available tracking methods can be divided into two main classes: feature-based and modelbased. The former approach focuses on tracking 2D features such as geometrical primitives (points, segments, circles,...), object contours, regions of interest, etc. The latter explicitly uses a model of the tracked objects. This can be either a 3D model or a 2D template of the object. This second class of methods usually provides a more robust solution. Indeed, the main advantage of the model-based methods is that the knowledge about the scene allows improving tracking robustness and performance, by being able to predict hidden movements of the object, detect partial occlusions and acts to reduce the effects of outliers. The challenge is to build algorithms that are fast and robust enough to meet our application requirements. Therefore, even if we still consider 2D feature tracking in some cases, our researches mainly focus on real-time 3D model-based tracking, since these approaches are very accurate, robust, and well adapted to any class of visual servoing schemes. Furthermore, they also meet the requirements of other classes of application, such as augmented reality.

# **3.3. SLAM**

Most of the applications involving mobile robotic systems (ground vehicles, aerial robots, automated submarines,...) require a reliable localization of the robot in its environment. A challenging problem is when neither the robot localization nor the map is known. Localization and mapping must then be considered concurrently. This problem is known as Simultaneous Localization And Mapping (SLAM). In this case, the robot moves from an unknown location in an unknown environment and proceeds to incrementally build up a navigation map of the environment, while simultaneously using this map to update its estimated position.

Nevertheless, solving the SLAM problem is not sufficient for guaranteeing an autonomous and safe navigation. The choice of the representation of the map is, of course, essential. The representation has to support the different levels of the navigation process: motion planning, motion execution and collision avoidance and, at the global level, the definition of an optimal strategy of displacement. The original formulation of the SLAM problem is purely metric (since it basically consists in estimating the Cartesian situations of the robot and a set of landmarks), and it does not involve complex representations of the environment. However, it is now well recognized that several complementary representations are needed to perform exploration, navigation, mapping, and control tasks successfully. We propose to use composite models of the environment that mix topological, metric, and grid-based representations. Each type of representation is well adapted to a particular aspect of autonomous navigation [7]: the metric model allows one to locate the robot precisely and plan Cartesian paths, the topological model captures the accessibility of different sites in the environment and allows a coarse localization, and finally the grid representation is useful to characterize the free space and design potential functions used for reactive obstacle avoidance. However, ensuring the consistency of these various representations during the robot exploration, and merging observations acquired from different viewpoints by several cooperative robots, are difficult problems. This is particularly true when different sensing modalities are involved. New studies to derive efficient algorithms for manipulating the hybrid representations (merging, updating, filtering...) while preserving their consistency are needed.

# 3.4. Scene Modeling and Understanding

Long-term mapping has received an increasing amount of attention during last years, largely motivated by the growing need to integrate robots into the real world wherein dynamic objects constantly change the appearance of the scene. A mobile robot evolving in such a dynamic world should not only be able to build a map of the observed environment at a specific moment, but also to maintain this map consistent over a long period of time.

It has to deal with dynamic changes that can cause the navigation process to fail. However updating the map is particularly challenging in large-scale environments. To identify changes, robots have to keep a memory of the previous states of the environment and the more dynamic it is, the higher will be the number of states to manage and the more computationally intensive will be the updating process. Mapping large-scale dynamic environments is then particularly difficult as the map size can be arbitrary large. Additionally, mapping many times the whole environment is not always possible or convenient and it is useful to take advantages of methods using only a small number of observations.

A recent trend in robotic mapping is to augment low-level maps with semantic interpretation of their content, which allows to improve the robot's environmental awareness through the use of high-level concepts. In mobile robot navigation, the so-called semantic maps have already been used to improve path planning methods, mainly by providing the robot with the ability to deal with human-understandable targets.

# 4. Application Domains

# 4.1. Application Domains

The natural applications of our research are obviously in robotics. In fact, researches undertaken in the Lagadic group can apply to all the fields of robotics implying a vision sensor. They are indeed conceived to be independent of the system considered (and the robot and the vision sensor can even be virtual for some applications).

Currently, we are mostly interested in using visual servoing for aerial and space application, micromanipulation, autonomous vehicle navigation in large urban environments or for disabled or elderly people.

We also address the field of medical robotics. The applications we consider turn around new functionalities of assistance to the clinician during a medical examination: visual servoing on echographic images, needle insertion, compensation of organ motion, etc.

Robotics is not the only possible application field to our researches. In the past, we were interested in applying visual servoing in computer animation, either for controlling the motions of virtual humanoids according to their pseudo-perception, or for controlling the point of view of visual restitution of an animation. In both cases, potential applications are in the field of virtual reality, for example for the design of video games, or virtual cinematography.

Applications also exist in computer vision and augmented reality. It is then a question of carrying out a virtual visual servoing for the 3D localization of a tool with respect to the vision sensor, or for the estimation of its 3D motion. This field of application is very promising, because it is in full rise for the realization of special effects in the multi-media field or for the design and the inspection of objects manufactured in the industrial world.

# 5. Highlights of the Year

# **5.1. Highlights of the Year**

- Claudio Pacchierotti has been elected Chair of the IEEE Technical Committee on Haptics for the term 2018-2020. He also published a review paper on the topic of wearable haptic devices for the hand [29].
- Julien Pettré will coordinate the H2020 ICT 25 EUropean Project CrowdBot starting from January 2018. The project gathers 5 academic partners UCL (UK), EPFL and ETHZ (Switzerland), RWTH (Germany) and Inria (France) as well as 2 industrial partners Locomotec GmbH (Germany) and SoftBank Robotics (France). The project will address the navigation of robots in crowded environment. While having robot moving in crowds can be of crucial importance (e.g., semi-autonomous wheelchairs), the project will design new robot navigation techniques that minimize the risk of negative impact raised by the presence of the robot (traffic perturbation, collision, etc.).

## 5.1.1. Awards

Lagadic was a member of the five finalist teams for the KUKA Innovation Award (https://www.kuka.com/en-de/press/events/kuka-innovation-award), together with the RIS group at LAAS (coordinator), the University of Siena, Italy, and the Seoul National University, South Korea. The goal was to address search and rescue operations in regions which are difficult to access or dangerous following disasters. For this, the team explored the collaboration between a quadrotor UAV and a KUKA lightweight arm for cooperative transportation and manipulation of rigid objects (e.g., long bards), with a final peg-in-hole task demonstrated live at the Hannover fair in April 2017.

# 6. New Software and Platforms

# 6.1. bib2html

Latex bibliography generator

KEYWORDS: LaTeX - Bibliography

FUNCTIONAL DESCRIPTION: The purpose of this software is to automatically produce html pages from BibTEX files, and to provide access to the BibTEX entries by several criteria: year of publication, category of publication, keywords, author name. Moreover cross-linking is generating between pages to provide an easy navigation through the pages without going back to the index.

- Contact: Éric Marchand
- URL: http://www.irisa.fr/lagadic/soft/bib2html/bib2html.html

# 6.2. DESlam

#### Dense Egocentric SLAM

**KEYWORDS: Deph Perception - Robotics - Localisation** 

FUNCTIONAL DESCRIPTION: This software proposes a full and self content solution to the dense Slam problem. Based on a generic RGB-D representation valid for various type of sensors (stereovision, multi-cameras, RGB-D sensors...), it provides a 3D textured representation of complex large indoor and outdoor environments and it allows localizing in real time (45Hz) a robot or a person carrying out a mobile camera.

- Participants: Andrew Ian Comport, Maxime Meilland and Patrick Rives
- Contact: Patrick Rives

# 6.3. HandiViz

Driving assistance of a wheelchair

KEYWORDS: Health - Persons attendant - Handicap FUNCTIONAL DESCRIPTION: The HandiViz software proposes a semi-autonomous navigation framework of a wheelchair relying on visual servoing.

It has been registered to the APP ("Agence de Protection des Programmes") as an INSA software (IDDN.FR.001.440021.000.S.P.2013.000.10000) and is under GPL license.

- Participants: François Pasteau and Marie Babel
- Contact: Marie Babel

# 6.4. Perception360

Robot vision and 3D mapping with omnidirectional RGB-D sensors. KEYWORDS: Depth Perception - Localization - 3D reconstruction - Realistic rendering - Sensors - Image registration - Robotics - Computer vision - 3D rendering FUNCTIONAL DESCRIPTION: This software is a collection of libraries and applications for robot vision and 3D mapping with omnidirectional RGB-D sensors or standard perspective cameras. This project provides the functionality to do image acquisition, semantic annotation, dense registration, localization and 3D mapping. The omnidirectional RGB-D sensors used within this project have been developed in Inria Sophia-Antipolis by the team LAGADIC.

• Contact: Patrick Rives

# **6.5. SINATRACK**

#### Model-based visual tracking of complex objects

**KEYWORDS:** Computer vision - Robotics

FUNCTIONAL DESCRIPTION: Sinatrack is a tracking software that allows the 3D localization (translation and rotation) of an object with respect to a monocular camera. It allows to consider object with complex shape. The underlying approach is a model-based tracking techniques. It has been developed for satellite localization and on-orbit service applications but is also suitable for augmented reality purpose.

- Participants: Antoine Guillaume Petit, Éric Marchand and François Chaumette
- Contact: Éric Marchand

# 6.6. UsTk

#### Ultrasound toolkit for medical robotics applications guided from ultrasound images

KEYWORDS: Echographic imagery - Image reconstruction - Medical robotics - Visual tracking - Visual servoing (VS)

FUNCTIONAL DESCRIPTION: UsTk, standing for Ultrasound Toolkit, is a cross-platform extension of ViSP software dedicated to two- and three-dimensional ultrasound image processing and visual servoing based on ultrasound images. Written in C++, UsTk architecture provides a core module that implements all the data structures at the heart of UsTk, a grabber module that allows to acquire ultrasound images from an Ultrasonix or a Sonosite device, a GUI module to display data, an IO module for providing functionalities to read/write data from a storage device, and a set of image processing modules to compute the confidence map, to track a needle, and to track an image template. All these modules could be used to control the motion of an ultrasound probe by ultrasound visual servoing.

- Participants: Alexandre Krupa, Marc Pouliquen, Fabien Spindler and Pierre Chatelain
- Partners: Université de Rennes 1 IRSTEA
- Contact: Alexandre Krupa
- URL: https://team.inria.fr/lagadic/

# 6.7. ViSP

#### Visual servoing platform

KEYWORDS: Augmented reality - Computer vision - Robotics - Visual servoing (VS)

SCIENTIFIC DESCRIPTION: Since 2005, we develop and release ViSP [1], an open source library available from https://visp.inria.fr. ViSP standing for Visual Servoing Platform allows prototyping and developing applications using visual tracking and visual servoing techniques at the heart of the Lagadic research. ViSP was designed to be independent from the hardware, to be simple to use, expandable and cross-platform. ViSP allows to design vision-based tasks for eye-in-hand and eye-to-hand visual servoing that contains the most classical visual features that are used in practice. It involves a large set of elementary positioning tasks with respect to various visual features (points, segments, straight lines, circles, spheres, cylinders, image moments, pose...) that can be combined together, and image processing algorithms that allow tracking of visual cues (dots, segments, ellipses...) or 3D model-based tracking of known objects or template tracking. Simulation capabilities are also available.
[1] E. Marchand, F. Spindler, F. Chaumette. ViSP for visual servoing: a generic software platform with a wide class of robot control skills. IEEE Robotics and Automation Magazine, Special Issue on "Software Packages for Vision-Based Control of Motion", P. Oh, D. Burschka (Eds.), 12(4):40-52, December 2005.

FUNCTIONAL DESCRIPTION: ViSP provides simple ways to integrate and validate new algorithms with already existing tools. It follows a module-based software engineering design where data types, algorithms, sensors, viewers and user interaction are made available. Written in C++, ViSP is based on open-source cross-platform libraries (such as OpenCV) and builds with CMake. Several platforms are supported, including OSX, iOS, Windows and Linux. ViSP online documentation allows to ease learning. More than 280 fully documented classes organized in 17 different modules, with more than 300 examples and 64 tutorials are proposed to the user. ViSP is released under a dual licensing model. It is open-source with a GNU GPLv2 license. A professional edition license that replaces GNU GPLv2 is also available.

- Participants: Aurélien Yol, Éric Marchand, Fabien Spindler, François Chaumette and Souriya Trinh
- Partner: Université de Rennes 1
- Contact: Fabien Spindler
- URL: http://visp.inria.fr

# 6.8. Platforms

# 6.8.1. Robot Vision Platform

Participant: Fabien Spindler [contact].

We exploit two industrial robotic systems built by Afma Robots in the nineties to validate our researches in visual servoing and active vision. The first one is a 6 DoF Gantry robot, the other one is a 4 DoF cylindrical robot (see Fig. 2.a). These robots are equipped with cameras. The Gantry robot also allows embedding grippers on its end-effector.

We are also using a haptic Virtuose 6D device from Haption company (see Fig. 2.b). This device is used as master device in many of our shared control activities (see Sections 9.3.1.3, 7.3.3, and 7.3.4).

Note that eight papers published by Lagadic in 2017 enclose results validated on this platform [35], [37], [15], [63], [58], [48], [51], [52].

# 6.8.2. Mobile Robots

Participants: Fabien Spindler [contact], Marie Babel, Patrick Rives.

#### 6.8.2.1. Indoor Mobile Robots

For fast prototyping of algorithms in perception, control and autonomous navigation, the team uses Hannibal in Sophia Antipolis, a cart-like platform built by Neobotix (see Fig. 3.a), and, in Rennes, a Pioneer 3DX from Adept (see Fig. 3.b). These platforms are equipped with various sensors needed for SLAM purposes, autonomous navigation, and sensor-based control.

Moreover, to validate the researches in personally assisted living topic (see Section 7.5.3), we have three electric wheelchairs in Rennes, one from Permobil, one from Sunrise and the last from YouQ (see Fig. 3.c). The control of the wheelchair is performed using a plug and play system between the joystick and the low level control of the wheelchair. Such a system lets us acquire the user intention through the joystick position and control the wheelchair by applying corrections to its motion. The wheelchairs have been fitted with cameras and ultrasound sensors to perform the required servoing for assisting handicapped people.

Note that five papers exploiting the indoors mobile robots were published this year [15], [30], [31], [53], [60].





Figure 1. This figure highlights ViSP main capabilities for visual tracking, visual servoing, and augmented reality that may benefit from computer vision algorithms. ViSP allows controlling specific platforms through hardware abstraction or in simulation. ViSP provides also bridges over other frameworks such as OpenCV and ROS. All these capabilities are cross-platform. Moreover, for easing the prototyping of applications, ViSP provides tools for image manipulation, mathematics, data plotting, camera calibration, and many other features. ViSP powerful API is fully documented and available on Github as an open source software under GPLv2 license.



(a) (b) Figure 2. a) Lagadic robotics platform for vision-based manipulation, b) Virtuose 6D haptic device

#### 6.8.2.2. Outdoor Vehicles

A camera rig has been developed in Sophia Antipolis. It can be fixed to a standard car (see Fig. 4), which is driven at a variable speed depending on the road/traffic conditions, with an average speed of 30 km/h and a maximum speed of 80 km/h. The sequences are recorded at a frame rate of 20 Hz, whith a synchronization of the six global shutter cameras of the stereo system, producing spherical images with a resolution of 2048x665 pixels (see Fig. 4). Such sequences are fused offline to obtain maps that can be used later for localization or for scene rendering (in a similar fashion to Google Street View) as shown in the video http://www-sop.inria. fr/members/Renato-Jose.Martins/iros15.html.

#### 6.8.3. Medical Robotic Platform

Participants: Marc Pouliquen, Fabien Spindler [contact], Alexandre Krupa.

This platform is composed by two 6 DoF Adept Viper arms (see Fig. 5.a). Ultrasound probes connected either to a SonoSite 180 Plus or an Ultrasonix SonixTouch imaging system can be mounted on a force torque sensor attached to each robot end-effector. The haptic Virtuose 6D device (see Fig. 2.b) can also be used within this platform.

This testbed is of primary interest for researches and experiments concerning ultrasound visual servoing applied to probe positioning, soft tissue tracking, elastography or robotic needle insertion tasks (see Section 7.3).

Note that seven papers published this year include experimental results obtained with this platform [56], [57], [72], [33], [19], [48], [37]

## 6.8.4. Humanoid Robots

Participants: Giovanni Claudio, Fabien Spindler [contact].

Romeo is a humanoid robot from SoftBank Robotics which is intended to be a genuine personal assistant and companion. Only the upper part of the body (trunk, arms, neck, head, eyes) is working. This research platform is used to validate our researches in visual servoing and visual tracking for object manipulation (see Fig. 6.a).



(a)



(b)



(c)

Figure 3. a) Hannibal platform, b) Pioneer P3-DX robot, c) wheelchairs from Permobil, Sunrise and YouQ.





Figure 4. Globeye stereo sensor and acquisition system.



(a)

(b)

Figure 5. a) Lagadic medical robotic platforms. On the right Viper S850 robot arm equipped with a SonixTouch 3D ultrasound probe. On the left Viper S650 equipped with a tool changer that allows to attach a classical camera or biopsy needles. b) Robotic setup for autonomous needle insertion by visual servoing.

Last year, this platform was extended with Pepper, another human-shaped robot designed by SoftBank Robotics to be a genuine day-to-day companion (see Fig. 6.b). It has 17 DoF mounted on a wheeled holonomic base and a set of sensors (cameras, laser, ultrasound, inertial, microphone) that makes this platform interesting for researches in vision-based manipulation, and visual navigation (see Section 7.5.1).

Note that two papers published this year include experimental results obtained with these platforms [13], [60].

#### 6.8.5. Unmanned Aerial Vehicles (UAVs)

Participants: Thomas Bellavoir, Pol Mordel, Paolo Robuffo Giordano [contact].

From 2014, Lagadic also started some activities involving perception and control for single and multiple quadrotor UAVs, especially thanks to a grant from "Rennes Métropole" (see Section 9.1.4) and the ANR project "SenseFly" (see Section 9.2.5). To this end, we purchased four quadrotors from Mikrokopter Gmbh, Germany (see Fig. 7.a), and one quadrotor from 3DRobotics, USA (see Fig. 7.b). The Mikrokopter quadrotors have been heavily customized by: (i) reprogramming from scratch the low-level attitude controller onboard the microcontroller of the quadrotors, (ii) equipping each quadrotor with an Odroid XU4 board (see Fig. 7.d) running Linux Ubuntu and the TeleKyb software (the middleware used for managing the experiment flows and the communication among the UAVs and the base station), and (iii) purchasing the Flea Color USB3 cameras together with the gimbal needed to mount them on the UAVs (see Fig. 7.c). The quadrotor group is used as robotic platforms for testing a number of single and multiple flight control schemes with a special attention on the use of onboard vision as main sensory modality.

This year four papers published enclose experimental results obtained with this platform [49], [50], [42], [62].

# 7. New Results

# 7.1. Visual Perception



(a) Figure 6. a) Romeo experimental platform, b) Pepper human-shaped robot



Figure 7. a) Quadrotor XL1 from Mikrokopter, b) Quadrotor Iris from 3DRobotics, c) Flea Color USB3 camera, d) Odroid XU4 board

(b)

## 7.1.1. Visual Tracking for Motion Capture

Participant: Eric Marchand.

This work is achieved in collaboration with Anatole Lécuyer (Inria Hybrid group) through the co-supervision of Guillaume Cortes Ph.D.

In the context of the development of new optical tracking devices, we propose an approach to greatly increase the tracking workspace of VR applications without adding new sensors [69]. Our approach relies on controlled cameras able to follow the tracked markers all around the VR workspace providing 6 DoF tracking data. We designed the proof-of-concept of such approach based on two consumer-grade cameras and a pan-tilt head. This approach has also been extended for the tracking of a drone in GPS denied environment [42].

We also achieved a short study related to the analysis of the 3D motion of head and hand in CAVE-based applications with the goal to optimize optical tracking sensors placement [43].

#### 7.1.2. Object 3D Tracking based on Depth Information and CAD Model

Participants: Agniva Sengupta, Eric Marchand, Alexandre Krupa.

In the context of the iProcess project (see Section 9.3.3.2), we started this year a new study related to pose estimation and tracking of a rigid object observed by a RGB-D camera. We developed a pose estimation approach based on depth information measurement and the use of a CAD model represented by a 3D tetrahedral mesh. The pose parameters are estimated through an iterative optimization process that minimizes the point-to-plane Euclidean distance between the point cloud observed by the RGB-D camera and the surface of the 3D mesh. Preliminary results obtained with simple objects constituted by a set of orthogonal planes showed good performance of this approach. However, the method failed for the case of complex objects that exhibit important curvature surfaces. In order to address this issue we are currently extending the approach to take into account also the RGB information in the optimization criterion.

#### 7.1.3. General Model-based Tracker

Participants: Souriya Trinh, Fabien Spindler, François Chaumette.

We have generalized the model-based tracker [2] available in ViSP [5] to integrate the depth information provided by a RGB-D sensor using the method described in the previous paragraph. It is now possible to fuse in the same optimization scheme measurements such as points of interest, edges, and depth, which allows to improve the robustness and accuracy of the tracker.

#### 7.1.4. 3D Localization for Airplane Landing

Participants: Noël Mériaux, Pierre-Marie Kerzerho, Patrick Rives, Eric Marchand, François Chaumette.

This study was realized in the scope of the ANR VisioLand project (see Section 9.2.2). In a first step, we have considered and adapted our model-based tracker [2] to localize the aircraft with respect to the airport surroundings. Satisfactory results have been obtained from real image sequences provided by Airbus. In a second step, we implemented a direct registration method based on dense vision-based tracking that allows localizing the on-board camera from a set of keyframe images corresponding to the landing trajectory. First experiments with simulated and real images have been carried on with promising results. This approach is particularly interesting at the beginning of the descent when the landing track is far away and not very observable in the image. In that sense, the direct registration method is strongly complementary with the model-based approach studied before.

## 7.1.5. Extrinsic Calibration of Multiple RGB-D Cameras

Participants: Eduardo Fernandez Moral, Patrick Rives.

In collaboration with Alejandro Perez-Yus from the University of Zaragoza, we developed a novel method to estimate the relative poses between RGB and depth cameras without the requirement of an overlapping field of view, thus providing flexibility to calibrate a variety of sensor configurations. This calibration problem is relevant to robotic applications which can benefit of using several cameras to increase the field of view. In

our approach, we extract and match lines of the scene in the RGB and depth cameras, and impose geometric constraints to find the relative poses between the sensors. In [31], an analysis of the observability properties of the problem is presented. We have validated our method in both synthetic and real scenarios with different camera configurations, demonstrating that our approach achieves good accuracy and is very simple to apply, in contrast with previous methods based on trajectory matching using visual odometry or SLAM.

#### 7.1.6. Scene Registration with Large Convergence Domain

Participants: Renato José Martins, Patrick Rives.

Image registration has been a major problem in computer vision over the past decades. It implies searching an image in a database of previously acquired images to find one (or several) that fulfill some degree of similarity, e.g. an image of the same scene from a similar viewpoint. This problem is interesting in mobile robotics for topological mapping, re-localization, loop closure and object identification. Scene registration can be seen as a generalization of the above problem where the representation to match is not necessarily defined by a single image (i.e. the information may come from different images and/or sensors), attempting to exploit all information available to pursue higher performance and flexibility. This problem is ubiquitous in robot localization and navigation. We propose a probabilistic framework to improve the accuracy and efficiency of a previous solution for structure registration based on planar representation [12]. The main idea is to explore the properties given by planar surfaces with co-visibility and their normals from two distinct viewpoints. We estimate, in two decoupled stages, the rotation and then the translation, both based on the normal vectors orientation and on the depth. These two stages are efficiently computed by using low resolution depth images and without any feature extraction/matching. In [53], we also analyze the limitations and observability of this approach, and its relationship to ICP point-to-plane. Notably, if the rotation is observable, at least five DoF can be estimated in the worst case. To demonstrate the effectiveness of the method, we evaluate the initialization technique in a set of challenging scenarios, comprising simulated spherical images from the Sponza Atrium model benchmark and real spherical indoor sequences.

#### 7.1.7. Scene Semantization based on Deep Learning Approach

Participants: Eduardo Fernandez Moral, Patrick Rives.

Semantic segmentation of images is an important problem for mobile robotics and autonomous driving because it offers basic information which can be used for complex reasoning and safe navigation. This problem constitutes a very active field of research, where the state-of-the-art evolves continuously with new strategies based on different kinds of deep neural networks for image segmentation and classification. RGB-D images are starting to be employed as well for the same purpose to exploit complimentary information from color and geometry. The team LAGADIC has explored several strategies to increase the performance and the accuracy of semantic segmentation from RGB-D images. We propose a multi-pipeline architecture to exploit effectively the complimentary information from RGB-D images and thus to improve the semantic segmentation results. The multi-pipeline architecture processes the color and depth layers in parallel, before concatenating their feature maps to produce the final semantic prediction. Our results are evaluated on public benchmark datasets to show the improved accuracy of the proposed architecture. [46] Though we address this problem in the context of urban images segmentation, our results can also be extended to other contexts, like indoor scenarios and domestic robotics.

Our research is partly motivated by the need of semantic segmentation solutions with better segmentation around contours. Besides, we note that one of the main issues when comparing different neural networks architectures is how to select an appropriate metric to evaluate their accuracy. We have studied several metrics for multi-class classification, and we propose a new metric which accounts for both global and contour accuracy in a simple formulation to overcome the weaknesses of previous metrics. This metric is based on the Jaccard index, and takes explicitly into account the distance to the border regions of the different classes, to encode jointly the rate of correctly labeled pixels and how homeomorphic is the segmentation to the real object boundaries. We also present a comparative analysis of our proposed metric and several commonly used metrics for semantic segmentation together with a statistical analysis of their correlation.

## 7.1.8. Online Localization and Mapping for UAVs

Participants: Muhammad Usman, Paolo Robuffo Giordano.

Localization and mapping in unknown environments is still an open problem, in particular for what concerns UAVs because of the typical limited memory and processing power available onboard. In order to provide our quadrotor UAVs with high autonomy, we started studying how to exploit onboard cameras for an accurate (but fast) localization and mapping in unknown indoor environments. We chose to base both processes on the newly available Semi-Direct Visual Odometry (SVO) library (http://rpg.ifi.uzh.ch/software) which has gained considerable attention over the last years in the robotics community. The idea is to exploit dense images (i.e., with little image pre-processing) for obtaining an incremental update of the camera pose which, when integrated over time, can provide the camera localization (pose) w.r.t. the initial frame. In order to reduce drifts during motion, a concurrent mapping thread is also used for comparing the current view with a set of keyframes (taken at regular steps during motion) which constitute a "map" of the environment. We have started porting the SVO library to our UAVs and the preliminary results showed good performance of the localization accuracy against the Vicon ground truth. We are now planning to close the loop and base the UAV flight on the reconstructed pose from the SVO algorithm.

#### 7.1.9. Reflectance and Illumination Estimation for Realistic Augmented Reality

Participants: Salma Jiddi, Eric Marchand.

A key factor for realistic Augmented Reality is a correct illumination simulation. This consists in estimating the characteristics of real light sources and use them to model virtual lighting. This year, we studied a novel method for recovering both 3D position and intensity of multiple light sources using detected cast shadows. Our algorithm has been successfully tested on a set of real scenes where virtual objects have visually coherent shadows [70].

#### 7.1.10. Optimal Active Sensing Control

Participants: Marco Cognetti, Paolo Salaris, Paolo Robuffo Giordano.

This study concerns the problem of active sensing control whose objective is to reduce the estimation uncertainty of an observer as much as possible by determining the inputs of the system that maximize the amount of information gathered by the few noisy outputs while at the same time reduce the negative effects of the process/actuation noise. The latter is far from being negligible for several robotic applications (a prominent example being aerial vehicles).

Last year, we extended a previous work [9] to the case where the observability property is not instantaneously guaranteed, and hence the optimal estimation strategy cannot be given in terms of the instantaneous velocity direction of the robot and consequently of the onboard sensors. These outcomes of this research have been presented in [61] for nonlinear differentially flat systems. This year, we have moved some steps forward in order to improve and generalize the work in [61]: first of all, we have replaced the Observability Gramian (OG) with the Constructibility Gramian (CG). Despite their similar form, they differ from the fact that the OG measures the information collected along the path about the initial state of the nonlinear system while the CG measures the one about the current/final state with which most robotics applications are more concerned. Second, we have overcome the limit of previous work [61] that only deals with the case where the OG and the CG are known in closed-form. We have also applied our method to the unicycle vehicle which is a more complex dynamic system than the one used in [61] and tested our machinery to the cases of self-calibration and environment reconstruction. Moreover, thanks to the arrival of Marco Cognetti in our group as Postdoc, we are currently working on the application of our method to a quadrotor UAV, which is a much more complex dynamic system, for which the CG is not known in closed-form. The ultimate goal is to test our new machinery in a real experiment with a quadrotor UAV. Finally, we have also worked on the problem of considering the process/actuation noise in the optimization algorithm. As the CG (or the OG) does not take into account the degrading effects on the information collected through the outputs of the process/actuation noise, we have proposed to directly maximize the smallest eigenvalue of the covariance matrix given by the Riccati differential equation of the EKF, used as estimation algorithm. The results of this approach have been submitted to ICRA 2018.

# 7.2. Sensor-based Robot Control

## 7.2.1. Determining Singularity Configurations in IBVS

Participant: François Chaumette.

This theoretical study has been achieved through an informal collaboration with Sébastien Briot and Philippe Martinet from LS2N in Nantes, France. It concerned the determination of the singularity configurations of image-based visual servoing using tools from the mechanical engineering community and the concept of "hidden" robot. In a first step, we have revisited the well-known case of using three image points as visual feature, and then solved the general case of n image points [16]. The case of three image straight lines has also been solved for the first time [17].

We have also designed a control scheme in order to avoid these singularities during the execution of a visual servoing scheme [38].

#### 7.2.2. Visual Servoing through Mirror Reflection

Participants: François Chaumette, Eric Marchand.

Apart the use of catadioptric cameras, only few visual servoing works exploit the use of mirror. Such a configuration is however interesting since it allows overpassing the limited camera field of view. Based on the known projection equations involved in such a system, we studied the theoretical background that allows the control of planar mirror for visual servoing in different configurations. Limitations intrinsic to such systems, such as the number of DoF actually controllable, have been studied. The case of point feature was considered in [51] and this has been extended to line in [52].

#### 7.2.3. Visual Servoing of Humanoid Robots

Participants: Giovanni Claudio, Fabien Spindler, François Chaumette.

This study is realized in the scope of the BPI Romeo 2 and H2020 Comanoid projects (see Sections 9.2.7 and 9.3.1.2).

We have designed the modeling of the visual features at the acceleration level to embed visual tasks and visual constraints in an existing Quadratic Programming controller [13]. Experimental results have been obtained on Romeo (see Section 6.8.4).

#### 7.2.4. Model Predictive Visual Servoing

Participants: Paolo Robuffo Giordano, François Chaumette.

This study was realized in collaboration with Pierre-Brice Wieber, from Bipop group at Inria Rhône Alpes, through the co-supervision of Nicolas Cazy's Ph.D.

Model Predictive Control (MPC) is a powerful control framework able to take explicitly into account the presence of constraints in the controlled system (e.g., actuator saturations, sensor limitations, and so on). In this study, we studied the possibility of using MPC for tackling one of the most classical constraints of visual servoing applications, that is, the possibility to lose tracking of features because of occlusions, limited camera field of view, or imperfect image processing/tracking. The MPC framework depends upon the possibility to predict the future evolution of the controlled system over some time horizon, for correcting the current state of the modeled system whenever new information (e.g., new measurements) become available. We have also explored the possibility of applying these ideas in a multi-robot collaboration scenario where a UAV with a downfacing camera (with limited field of view) needs to provide localization services to a team of ground robots [41].

# 7.2.5. Model Predictive Control for Visual Servoing of a UAV

Participants: Bryan Penin, François Chaumette, Paolo Robuffo Giordano.

Visual servoing is a well-known class of techniques meant to control the pose of a robot from visual input by considering an error function directly defined in the image (sensor) space. These techniques are particularly appealing since they do not require, in general, a full state reconstruction, thus granting more robustness and lower computational loads. However, because of the quadrotor underactuation and inherent sensor limitations (mainly limited camera field of view), extending the classical visual servoing framework to the quadrotor flight control is not straightforward. For instance, for realizing a horizontal displacement the quadrotor needs to tilt in the desired direction. This tilting, however, will cause any downlooking camera to point in the opposite direction with, e.g., possible loss of feature tracking because of the limited camera field of view.

In order to cope with these difficulties and achieve a high-performance visual servoing of quadrotor UAVs, we chose to rely on MPC for explicitly dealing with this kind of constraints during flight. We have recently considered the problem of controlling in minimum-time a quadrotor UAV equipped with a downlooking camera that needs to reach a desired pose w.r.t. a target on the ground from visual input. The control problem is solved by an online replanning strategy that is able to generate (at camera rate) minimum-time trajectories towards the final pose while coping with actuation constraints (limited propeller thrusts) and sensing constraints (target always in the camera fov). By exploiting the camera images during motion, the replanning strategy is able to adjust online the optimal trajectory and, thus, be robust against unmodeled effects and other disturbances (which can be typically expected on a quadrotor flying aggressively). The approach has been validated via numerical simulations in [59]. We are now working towards an experimental validation, as well as novel algorithmic extensions allowing for the possibility of temporarily losing sight of the target object for relaxing the visibility constraint (and, thus, gain in maneuverability).

#### 7.2.6. UAVs in Physical Interaction with the Environment

Participants: Quentin Delamare, Paolo Robuffo Giordano.

Most research in UAVs deals with either contact-free cases (the UAVs must avoid any contact with the environment), or in "static" contact cases (the UAVs need to exert some forces on the environment in quasistatic conditions, reminiscent of what has been done with manipulator arms). Inspired by the vast literature on robot locomotion (from, e.g., the humanoid community), in this research topic we aim at exploiting the contact with the environment for *helping* a UAV maneuvering in the environment, in the same spirit in which we humans (and, supposedly, humanoid robots) use our legs and arms when navigating in cluttered environments for helping in keeping balance, or perform maneuvers that would be, otherwise, impossible.

As an initial case study, we have considered a planar UAV equipped with a 1 DoF actuated arm capable of hooking at some pivots in the environment. This UAV (named MonkeyRotor) needs to "jump" from one pivot to the next one by exploiting the forces exchanged with the environment (the pivot) and its own actuation system (the propellers). This study considers the full dynamics in both cases (hooked, free-flying), proposes an optimization problem for finding optimal trajectories from an initial hooked configuration to the next one, and validates the approach in simulation. We are now working towards a physical realization of a first prototype. This activity is done in cooperation with LAAS-CNRS (Dr. Antonio Franchi who is co-supervising Quentin Delamare).

#### 7.2.7. Visual Servoing for Steering Simulation Agents

Participants: Axel Lopez Gandia, Eric Marchand, François Chaumette, Julien Pettré.

Steering is one of the basic functionality of any character animation system. It provides characters with the ability to locally move in the environment so as to achieve basic navigation tasks, such as reaching a goal, avoiding a collision with an obstacles, etc. This problem has been explored in various contexts (e.g., motion planning, autonomous characters or crowd simulation). It turned out that this component plays an important role on the quality of character animation and received a lot of attention. Many important steps have been taken to improve steering techniques: potential fields, sets of attractive and repulsive forces, linear programming in the velocity space, local optimization of navigation functions, etc. Each new category of approach leads to characters close to forming realistic trajectories when achieving navigation.

Nevertheless, all these techniques remain quite far from the way real humans form their locomotion trajectory, because they are all based on kinematics and geometry. Humans obviously do not solve geometrical problems of this nature while moving in their environment but control their motion from perceptual features, and more especially visual features they perceive from the environment. For simulating more accurately the perception-action loop used by humans to navigate in their environment, we developed a technique which provides characters with vision capabilities, by equipping them with a virtual retina on which we project information about their surroundings. In a first version, we projected information about the relative motion of objects around them, allowing characters to estimate the risk of collision they face, and to move so as to minimize this risk [21]. More recently, we projected a purely visual information, and we established the relations that exist between the visual features characters perceive and the motion they perform. This way, we are able to steer characters so as their visual flow satisfies some conditions, allowing them for example to reach a goal while avoiding surrounding obstacles, could they be static or moving.

## 7.2.8. Direct Visual Servoing

#### Participants: Quentin Bateux, Eric Marchand.

We have proposed a deep neural network-based method to perform high-precision, robust and real-time 6 DoF visual servoing [63]. We studied how to create a dataset simulating various perturbations (occlusions and lighting conditions) from a single real-world image of the scene. A convolutional neural network is fine-tuned using this dataset to estimate the relative pose between two images of the same scene. The output of the network is then employed in a visual servoing control scheme. The method converges robustly even in difficult real-world settings with strong lighting variations and occlusions.

# 7.3. Medical Robotics

## 7.3.1. Visual Servoing using Wavelet and Shearlet Transforms

Participants: Lesley-Ann Duflot, Alexandre Krupa.

In collaboration with Femto-ST lab in Besançon and the Research Group on Computational Data Analysis at Universitat Bremen, we developed a new generation of direct visual servoing methods in which the signal control inputs are the coefficients of a multiscale image representation. In particular, we considered the use of multiscale image representations that are based on discrete wavelet and shearlet transforms. We succeeded to derive an analytical formulation of the interaction matrix related to the wavelet and shearlet coefficients and experimentally demonstrated the performances of the proposed visual servoing approaches. We also considered this control framework in the design of a medical application which consists in automatically moving a biological sample carried by a parallel micro-robotic platform using Optical Coherence Tomography (OCT) as visual feedback. The objective is to automatically retrieve the region of the sample that corresponds to an initial optical biopsy for diagnosis purpose. First results obtained with a 3 DoF eye-to-hand visual servoing demonstrated the feasibility to use the wavelet coefficients of the OCT image as input of the control law.

# 7.3.2. 3D Steering of Flexible Needle by Ultrasound Visual Servoing

Participants: Jason Chevrie, Marie Babel, Alexandre Krupa.

We pursued our work on 3D steering of a flexible needle using ultrasound visual servoing [11]. This year, in collaboration with the Surgical Robotics Laboratory of the University of Twente, we developed a method to control a 2 DoF needle insertion device attached to the end-effector of a 6-DoF robotic arm in order to automatically insert a flexible needle toward a spherical target embedded in a moving biological tissue (bovine liver). We proposed a method that uses both base manipulation control and tip-based control while compensating the tissue motion to avoid lateral tearing. The visual feedback provided by the ultrasound probe was used to track the target and an electromagnetic tracker attached inside the needle was used to locate its tip. In this study, the motion compensation of the moving tissue was performed by minimizing the interaction force measured at the base of the needle insertion device. In our approach we used the generic task control framework to fuse the needle targeting and motion compensation tasks into a single control law. First experimental ex-vivo results demonstrated the efficiency of the proposed control to reach a target in moving biological tissue.

# 7.3.3. Robotic Assistance for Ultrasound Elastography by Visual Servoing, Force Control and Teleoperation

Participants: Pedro Alfonso Patlan Rosales, Alexandre Krupa.

This work concerns the development of a robotic assistant system for quantitative ultrasound elastography. This imaging modality provides the elastic parameters of a tissue which are commonly related with a certain pathology. It is performed by applying continuous stress variation on the tissue in order to estimate a strain map. Usually, this stress variation is performed manually by the user through the manipulation of the ultrasound probe and it results therefore in an user-dependent quality of the strain map. To improve the ultrasound elastography imaging and provide quantitative measurement, we developed an assistant robotic palpation system that automatically moves a 2D ultrasound probe for optimizing ultrasound elastography [72]. This year we extended our previous robotic palpation system in order to perform 3D elastography and allow the user to teleoperate the probe orientation through a haptic device [56]. This extension is based on the use of a 3D ultrasound probe held by a 6 DoF robotic arm and the design of a new control law based on the task control framework that simultaneously performs three tasks: i) autonomous palpation by force control of the tissue required for the strain map estimation, ii) probe lateral alignment on a stiff target of interest for optimizing its visibility by visual servoing and iii) teleoperation of the probe orientation by the user for exploration purpose. Recently, we also proposed a solution that allows the estimation of the strain map of a moving tissue that is subject to physiological motion [57]. It is based on the combination of a non-rigid motion tracking of the tissue of interest in the ultrasound image and an automatic 6 DoF compensation of the perturbation motion by visual servoing using dense ultrasound information.

#### 7.3.4. Haptic Guidance of a Biopsy Needle

Participants: Hadrien Gurnel, Alexandre Krupa.

We started a new study in collaboration with Maud Marchal (Inria Hybrid group) related to the assistance of manual needle steering for biopsies or therapy purposes (see Section 9.1.7). Instead of automatically inserting the needle by a robotic arm as we did in other works, our objective is to develop a solution that provides haptic cue feedback to the clinician that helps him during its manual gesture. The haptic cue feedback will be provided by a haptic device holding the needle. This year we developed a software tool that simulates and visualizes the interaction of a virtual needle with a deformable virtual organ. This organ is represented by a 3D mesh and a mass-spring-damper model was considered to simulate its deformation due to the needle insertion motion. The development of this software was based on our libraries UsTk and ViSP and the external library VTK (Visualization Toolkit). We also interfaced to this simulator our Haption Virtuose 6D haptic device to allow the user to teleoperate the virtual needle and to feel the force applied by the needle on the virtual tissue. This simulator will constitute an important tool for our future development of dynamic haptic guides before testing them in a real experimental setup.

# 7.4. Teleoperation

#### 7.4.1. Shared Control for Remote Manipulation

Participants: Firas Abi Farraj, Paolo Robuffo Giordano.

This work concerns our activities in the context of the RoMaNS H2020 project (see Section 9.3.1.3). Our main goal is to allow a human operator to be interfaced in an intuitive way with a two-arm system, one arm carrying a gripper (for grasping an object), and the other one carrying a camera for looking at the scene (gripper + object) and providing the needed visual feedback. The operator should be allowed to control the two-arm system in an easy way for letting the gripper approaching the target object, and she/he should also receive force cues informative of how feasible her/his commands are w.r.t. the constraints of the system (e.g., joint limits, singularities, limited camera fov, and so on).

We have started working on this topic by proposing a shared control architecture in which the operator could provide instantaneous velocity commands along four suitable task-space directions not interfering with the main task of keeping the gripper aligned towards the target object (this main task was automatically regulated). The operator was also receiving force cues informative of how much her/his commands were conflicting with the system constraints, in our case joint limits of both manipulators. Finally, the camera was always moving so as to keep both the gripper and the target object at two fixed locations on the image plane. Recently, we have extended this framework in several directions:

- in a first extension, the existing instantaneous interface has been improved towards an "integral" approach in which the user can command parts of the future manipulator trajectory, while the autonomy makes sure that no constraint is violated (in this case we considered, again, joint limits and singularities, as well as a more realistic vision constraint for keeping the gripper and the object always in visibility and not overlapping). This shared control algorithm was validated in simulation in [58]. We are currently completing a full implementation on our dual-arm system (the two Viper robots);
- 2. second, we have studied how to integrate learning from demonstration into our framework by first using learning techniques for extracting statistical regularities of "expert users" executing successful trajectories for the gripper towards the target object. Then, these learned trajectories were used for generating force cues able to guide novice users during their teleoperation task by the "hands" of the expert users who demonstrated the trajectories in the first place [37];
- 3. third, we have considered a grasping scenario in which a post-grasp task is specified (e.g., the grasped object needs to follow a predefined trajectory): in this scenario, the operator (supported by the robot autonomy) needs to decide where to best grasp in order to then execute the desired post-grasp action. However, different grasping poses will result in easier/harder execution by the robot because of any possible constraint (e.g., joint limits and singularities). Since awareness of these constraints is hard for any operator, in this case the autonomy component cues the operator with a force feedback indicating the best grasp pose w.r.t. the existing constraints and post-grasp task. The operator has still control over where to grasp, but she/he is guided by the force feedback into more feasible grasp poses than what she/he could have guessed without any feedback [48];
- 4. finally, we have considered the task of assisting an operator in control of a UAV which is mapping a remote environment with an onboard camera. In this scenario the operator can control the UAV motion during the mapping task. However, as in any estimation problem, different motions will result less/more optimal w.r.t. the scene estimation task: therefore, a force feedback is produced in order to assist the operator in selecting the UAV motion (in particular, its linear velocity) that also results optimal for the sake of facilitating the scene estimation process. The results have been validated with numerical simulations in a realistic environment [39].

#### 7.4.2. Wearable haptics

Participants: Marco Aggravi, Claudio Pacchierotti.

Kinesthetic haptic feedback is used in robotic teleoperation to provide the human operator with force information about the status of the slave robots and their interaction with the remote environment. Although kineshetic feedback has been proven to enhance the performance of teleoperation systems, it still shows several limitations, including its negative effect on the safety and stability of such systems, or the limited workspace, available DoF, high cost, and complexity of kinesthetic interfaces. In this respect, wearable haptics is gaining great attention. Safe, compact, unobtrusive, inexpensive, easy-to-wear, and lightweight haptic devices enable researchers to provide compelling touch sensations to multiple parts of the body, significantly increasing the applicability of haptics in many fields, such as robotics, rehabilitation, gaming, and immersive systems.

In this respect, our objective has been to study, design, and evaluate novel wearable haptic interfaces for the control of remote robotic systems as well as interacting with virtual immersive environments.

We have started by working on a multi-point wearable feedback solution for robotic manipulators operating in a cluttered environment [40]. The slave system is composed of an anthropomorphic soft robotic hand attached to a 6-axis force-torque sensor, which is in turn fixed to a 6-DoF robotic arm. The master system is composed of a Leap Motion controller and two wearable vibrotactile armbands, worn on the forearm and upper arm. The Leap Motion tracks the user's hand pose to control the pose of the manipulator and the grasping configuration of the robotic hand. The armband on the forearm conveys information about collisions of the slave hand/wrist system (green patch to green armband, see Fig. 8), whereas the armband on the upper arm conveys information about collisions of the slave arm (orange patch to orange armband). The amplitude of the vibrotactile feedback relayed by the armbands is proportional to the interaction force of the collision. A camera mounted near the manipulator's end-effector enables the operator to see the environment in front of the robotic hand. To validate our system, we carried out a human subjects telemanipulation experiment in a cluttered scenario. Twelve participants were asked to control the motion of the robotic manipulator to grasp an object hidden between debris of various shapes and stiffnesses. Haptic feedback provided by our wearable devices significantly improved the performance of the considered telemanipulation tasks. Finally, all subjects but one preferred conditions with wearable haptic feedback.



Figure 8. Haptic-enabled teleoperation system. We used two vibrotactile wearable devices to provide multi-point haptic feedback about collisions of the slave robot with the remote environment [40].

We have also used wearable haptics for guidance [20]. In this context, haptic feedback is not used to provide information about a force exerted by the salve robot in the remote environment, but it provides guidance cues about a predetermined trajectory to follow. Toward this, we developed a novel wearable device for the forearm. Four cylindrical rotating end effectors, located on the user's forearm, can generate skin stretch at the ulnar, radial, palmar, and dorsal sides of the arm. When all the end effectors rotate in the same direction, the cutaneous device is able to provide cues about a desired pronation/supination of the forearm. On the other hand, when two opposite end effectors rotate in opposite directions, the device is able to provide cutaneous cues about a desired translation of the forearm. Combining these two stimuli, we can provide both rotation and translation guidance. To evaluate the effectiveness of our device in providing navigation information, we carried out two experiments of haptic navigation. In the first one, subjects were asked to translate and rotate the forearm toward a target position and orientation, respectively. In the second experiment, subjects were asked to control a 6-DoF robotic manipulator to grasp and lift a target object. Haptic feedback provided by our wearable device improved the performance of both experiments with respect to providing no haptic feedback.

Moreover, it showed similar performance with respect to sensory substitution via visual feedback, without overloading the visual channel.

Finally, we also used wearable haptics for immersive virtual and augmented reality experiences, mainly addressing tasks related to entertainment and industrial training. In these case, we used wearable devices for the fingertips able to provide pressure and skin stretch sensations [24]. This article has also been featured in the News section of Science Magazine.

We also presented a review paper on the topic of wearable haptic devices for the hand [29].

# 7.5. Navigation of Mobile Robots

#### 7.5.1. Visual Navigation from an Image Memory

Participants: Paolo Robuffo Giordano, François Chaumette.

This study achieved during Suman Raj Bista's Ph.D. was concerned with visual autonomous navigation in indoor environments. As in our previous works concerning navigation outdoors [4], the approach is based on a topological localization of the current image with respect to a set of keyframe images, but the visual features used for this localization as well as for the visual servoing are not composed of points of interest only, but on a combination of points of interest and straight lines since they are more common indoors [60]. Satisfactory experimental results have been obtained using the Pioneer mobile robot (see Section 6.8.2) and Pepper (See Section 6.8.4).

#### 7.5.2. Robot-Human Interactions during Locomotion

#### Participant: Julien Pettré.

In collaboration with the Gepetto team of Laas in Toulouse and the Mimetic group in Rennes, we have studied how humans avoid collision with a robot. Understanding how humans achieve such avoidance is crucial to better anticipate humans' reactions to the presence of a robot and to control the robot to adapt its trajectory accordingly. It is generally assumed that humans avoid a robot just like they avoid another human. Last year, we brought the empirical evidence that humans actually set a specific strategy to avoid robots: they showed a preference to give way to a robot [36]. However, the robot was passive, i.e., not reacting to the presence of participants. This year, we studied interactions between humans and reactive robot, performing avoidance maneuvers to avoid collisions. Our conclusions are that, in such situations of human-robot interactions, human behave again as during human-human avoidance interactions. Again, this study provides useful guidelines about the design of robot control techniques.

#### 7.5.3. Semi-Autonomous Control of a Wheelchair for Navigation Assistance

Participants: Louise Devigne, Marie Babel.

In order to improve the access to mobility for people with disabilities, we have previously designed a semiautonomous assistive wheelchair system which progressively corrects the trajectory as the user manually drives the wheelchair and smoothly avoids obstacles. Within the frame of ISI4NAVE associated team (see Section 9.4.1.2), we investigated probabilistic blending approaches which take into account uncertainty in the interaction [45]. We also designed a shared-control curb-following solution for outdoor assisted power wheelchair navigation. Once a curb is detected, user input is blended with constraints deduced from the distance from sensors to the detected curb. This provides an intuitive shared control scheme capable of assisting the user while needed i.e. while approaching a curb. Preliminary validation tests of the robotic system were conducted within the PAMELA facility. Developing and testing such systems for wheelchair driving assistance requires a significant amount of material resources and clinician time. With Virtual Reality technology, prototypes can be developed and tested in a risk-free and highly flexible Virtual Environment before equipping and testing a physical prototype. Additionally, users can "virtually" test and train more easily during the development process. We then designed a power wheelchair driving simulator allowing the user to navigate with a standard wheelchair in an immersive 3D Virtual Environment. In order to validate the framework including the driving assistance solution, we performed tests on the Immersia platform (Inria Hybrid team) with able-bodied participants and we have shown that the simulator it generates a good sense of presence and requires rather low cognitive effort from users [44].

# 7.5.4. Wheelchair Kinematics and Dynamics Modeling for Shared Control

Participants: Aline Baudry, Marie Babel.

The driving experience of an electric powered wheelchair can be disturbed by unpleasant dynamic effects of the caster wheels, particularly during maneuvers in narrow rooms and direction changes. In order to prevent their nasty behaviour, we propose to model caster wheel kinematics and dynamics in order to implement a control law for a semi-autonomous assistance to maneuver in narrow environments. We conducted a preliminary study that has been achieved for our three types of wheelchair, each presenting different kinematic behaviors: front caster type, rear caster type and mid-wheel drive (see Figure 3.c). Transfer functions for each of these configurations have been identified. We achieved to design a parametric transfer function of the caster's behavior regarding to the initial orientation, wheelchair's velocity and user mass, in order to develop a sensorless maneuver control law.

#### 7.5.5. Wheelchair Autonomous Navigation for Fall Prevention

Participants: Solenne Fortun, Marie Babel.

The Prisme project (see Section 9.1.8) is devoted to fall prevention and detection of inpatients with disabilities. For wheelchair users, falls typically occur during transfer between the bed and the wheelchair and are mainly due to a bad positioning of the wheelchair. In this context, the Prisme project addresses both fall prevention and detection issues by means of a collaborative sensing framework. Ultrasonic sensors are embedded onto both a robotized wheelchair and a medical bed. The measured signals are used to detect fall and to automatically drive the wheelchair near the bed at an optimal position determined by occupational therapists. We first designed a detection solution based on a multiple echoes technique that enhances the system perception abilities. This augmented perception system is planned to be used for wheelchair navigation as well as fall detection.

#### 7.5.6. Robotic Platform for Assistance to People with Reduce Mobility

Participants: Dayana Hassan, Paolo Salaris, Patrick Rives.

The main objective of this work is to develop, in collaboration with AXYN Robotics (see Section 8.2.4), an intelligent vehicle to help elderly or persons with reduced mobility to move safely within a retirement home, an hospital or other much more crowded and dynamic environments. First of all, the vehicle has to be able to move within the environment while at the same time update the current map as accurately as possible. Once the map of the environment is available, the robot has to be able to plan the trajectory and reach a given destination. The robot should also follow a person taking into account social behaviors or bring towards a given destination, e.g. the canteen, making sure that an elderly person, affected e.g. by Alzheimer's disease, follows the robot. The robot should also work as an intelligent walker and help people in case of falling. In all these cases, it is very important to include humans (i.e. his/her model, his/her behaviors, his/her intentions etc.) within the study in order to develop adaptable human-aware path planning and control strategies. During this first year, the problem of following a person has been studied, starting from the literature, in order to find a suitable control scheme that merges feedback control laws, aimed at reactively cope with neighborhood environment events and feedforward ones, mainly intended to take into account the intentions of the person to follow, also including social behaviors.

# 7.6. Multi-robot and Crowd Motion Control

# 7.6.1. Rigidity-based Methods for Formation Control

Participants: Fabrizio Schiano, Paolo Robuffo Giordano.

Most multi-robot applications must rely on *relative sensing* among the robot pairs (rather than absolute/external sensing such as, e.g., GPS). For these systems, the concept of *rigidity* provides the correct framework for defining an appropriate sensing and communication topology architecture. Rigidity is a combinatorial theory for characterizing the "stiffness" or "flexibility" of structures formed by rigid bodies connected by flexible linkages or hinges. In a broader context, rigidity turns out to be an important architectural property of many multi-agent systems when a common inertial reference frame is unavailable. Applications that rely on sensor fusion for localization, exploration, mapping and cooperative tracking of a target, all can benefit from notions in rigidity theory. The concept of rigidity, therefore, provides the theoretical foundation for approaching decentralized solutions to the aforementioned problems using distance measurement sensors, and thus establishing an appropriate framework for relating system level architectural requirements to the sensing and communication capabilities of the system.

In our previous works we have addressed the problem of coordinating a team of quadrotor UAVs equipped with onboard cameras from which one could extract "relative bearings" (unit vectors in 3D) w.r.t. the neighboring UAVs in visibility. This problem is known as bearing-based formation control and localization. The basic assumption, however, was to always have a bearing rigid graph which may easily conflict with any sensing/communication constraint (measurements/edges can be lost whenever, e.g., a UAV leaves the camera fov, or it is occluded by another UAV/obstacle). In [62] we have then tackled the problem of "bearing rigidity maintenance" by studying how to formalize the problem of maintaining bearing rigidity over time despite possible sensing/communication constraints (min/max range, limited camera fov and occlusions in the reported work). Thanks to a suitable weighing machinery, we could define a "bearing rigidity eigenvalue" as a suitable metric for quantifying the degree of rigidity in the interaction graph, and then we could propose a gradient-based controller able to maintain the rigidity eigenvalue always positive (and, thus, guarantee bearing rigidity maintenance). The approach has been validated by experiments run on 5 quadrotor UAVs.

## 7.6.2. Cooperative Localization using Interval Analysis

Participants: Ide Flore Kenmogne Fokam, Vincent Drevelle.

In the context of multi-robot fleets, cooperative localization consists in gaining better position estimate through measurements and data exchange with neighboring robots. Positioning integrity (i.e., providing reliable position uncertainty information) is also a key point for mission-critical tasks, like collision avoidance. The goal of this work is to compute position uncertainty volumes for each robot of the fleet, using a decentralized method (i.e., using only local communication with the neighbors). The problem is addressed in a bounded-error framework, with interval analysis and constraint propagation methods. These methods enable to provide guaranteed position error bounds, assuming bounded-error measurements. They are not affected by over-convergence due to data incest, which makes them a well sound framework for decentralized estimation. Results have been obtained for image-based localization of a single UAV, enabling to characterize the pose uncertainty domain from measurements uncertainties [50], and also fusion with onboard proprioceptive sensors [49]. Extension to cooperative localization in a multi-UAV fleet has been studied in the two-robot case and continues as an ongoing work.

# 8. Bilateral Contracts and Grants with Industry

# 8.1. Bilateral Contracts with Industry

## 8.1.1. Robocortex

Participants: Souriya Trinh, Fabien Spindler, François Chaumette.

This contract with the Inria Robocortex start up in Sophia-Antipolis started in September 2016. It is devoted to provide our expertise in visual tracking for an application specified by Dassault Aviation.

#### 8.1.2. ABB

Participants: Souriya Trinh, Fabien Spindler, François Chaumette.

no Inria Rennes 12597, duration: 8 months.

This contract with ABB in Barcelona started in September 2017. It is devoted to provide our expertise in visual tracking and visual servoing for an industrial application.

## 8.1.3. IRT b<>com

Participants: Hadrien Gurnel, Fabien Spindler, Alexandre Krupa.

no Inria Rennes 11774, duration: 36 months.

This contract started in October 2016 and concerns the leasing to IRT b <> com of two modules of the Lagadic medical robotic platform. Each module is rent 40 days during a 3-year period in the context of the IRT <math>b <> com NeedleWare project (see Section 9.1.7).

# 8.2. Bilateral Grants with Industry

#### 8.2.1. Technicolor

Participants: Salma Jiddi, Eric Marchand.

no Univ. Rennes 1 15CC310-02D, duration: 36 months.

This project funded by Technicolor started in October 2015. It supports Salma Jiddi's Ph.D. about augmented reality (see Section 7.1.9).

# 8.2.2. Realyz

Participant: Eric Marchand.

no Inria Rennes 10822, duration: 36 months.

This project funded by Realyz started in October 2015. It is achieved in cooperation with Anatole Lécuyer from Hybrid group at Irisa and Inria Rennes-Bretagne Atlantique to support Guillaume Cortes Ph.D. about motion tracking in virtual reality.

## 8.2.3. Pôle Saint Hélier

Participants: Louise Devigne, Marie Babel.

no Insa Rennes 2015/0890, duration: 36 months.

This project started in November 2015. It addresses the following two issues. First, the idea is to design a low-cost indoor / outdoor efficient obstacle avoidance system that respects the user intention, and does not alter user perception. This involves embedding innovative sensors to tackle the outdoor wheelchair navigation problem. The second objective is to take advantage of the proposed assistive tool to enhance the user Quality of Experience by means of biofeedback as well as the understanding of the evolution of the pathology.

# 8.2.4. Axyn

Participants: Dayana Hassan, Paolo Salaris, Patrick Rives.

no Inria Sophia 10874-1, duration: 36 months.

The objective of this project that started in November 2016 is to explore new methodologies for the interaction between humans and robots, autonomous navigation and mapping and to transfer the results obtained on the robotic platform developed by AXYN for assisting disabled/elderly people at home or in hospital structures. Cost limits, good accessibility to aged people, robustness and safety related to the applications are at the heart of the project. This contract (ANRT-CIFRE) support Dayana Hassan's Ph.D (see Section 7.5.6).

# 9. Partnerships and Cooperations

# 9.1. Regional Initiatives

#### 9.1.1. ARED DeSweep

Participants: Lesley-Ann Duflot, Alexandre Krupa.

no Inria Rennes 8033, duration: 36 months.

This project funded by the Brittany council started in October 2014. It supports in part Lesley-Ann Duflot's Ph.D. about visual servoing based on the shearlet transform (see Section 7.3.1).

## 9.1.2. ARED Locoflot

Participants: Ide Flore Kenmogne Fokam, Vincent Drevelle, Eric Marchand.

no Inria Rennes 9944, duration: 36 months.

This project funded by the Brittany council started in October 2015. It supports in part Ide Flore Kenmogne Fokam's Ph.D. about cooperative localization in multi-robot fleets using interval analysis (see Section 7.6.2).

## 9.1.3. ARED Mod4Nav

Participants: Aline Baudry, Marie Babel.

no INSA Rennes 2016/01, duration: 36 months.

This project funded by the Brittany council started in October 2016. It supports in part Aline Baudry's Ph.D. about wheelchair modeling.

# 9.1.4. "Equipement mi-lourd Rennes Métropole"

Participant: Paolo Robuffo Giordano.

no CNRS Rennes 14C0481, duration: 36 months.

This grant from "Rennes Métropole" has been obtained in June 2014 and supported the activities related to the use of drones (quadrotor UAVs). The platform described in Section 6.8.5 has been purchased in part thanks to this grant.

# 9.1.5. "Allocation d'installation scientifique"

Participant: Claudio Pacchierotti.

no CNRS Rennes 17C0487, duration: 36 months.

This grant from "Rennes Métropole" has been obtained in July 2017 and supported the activities related to the teleoperation of drones (quadrotor UAVs) using wearable haptics interfaces.

# 9.1.6. IRT Jules Verne Mascot

Participant: François Chaumette.

no Inria Rennes 10361, duration: 36 months.

This project started in October 2015. It is managed by IRT Jules Verne in Nantes and achieved in cooperation with LS2N, Airbus, Renault, Faurecia and Alsthom. Its goal is to perform screwing for various industrial applications.

## 9.1.7. IRT b<>com NeedleWare

Participants: Hadrien Gurnel, Alexandre Krupa.

no Inria Rennes 9072, duration: 36 months.

This project started in October 2016. It supports Hadrien Gurnel's Ph.D. about the study of a shared control strategy fusing haptic and ultrasound visual control for assisting manual steering of needles for biopsy or therapy purposes in a synergetic way (see Section 7.3.4).

# 9.1.8. Prisme

Participants: Solenne Fortun, Marie Babel.

no Insa Rennes 9072, duration: 24 months.

This project started in January 2017 and is supported by Brittany region/BPI. This project aims at designing a fall prevention strategy based on the sensing collaboration of a smart wheelchair and a smart medical bed. Fall detection and automatic positioning of the wheelchair next to the bed issues are planned to be addressed (see Section 7.5.5).

# 9.2. National Initiatives

#### 9.2.1. France Life Imaging WP3-FLI ANFEET

Participant: Alexandre Krupa.

#### duration: 24 months.

This project started in January 2016. Its objective is to initiate collaborative research with the ICube laboratory (Strasbourg) on the control and supervision of flexible endoscopes in the digestive tube using ultrasound images.

## 9.2.2. ANR Contint Visioland

Participants: Noël Mériaux, Pierre-Marie Kerzerho, Patrick Rives, François Chaumette.

no Inria Rennes 8304, duration: 48 months.

This project ended in October 2017. It involved a consortium managed by Onera in Toulouse with Airbus, Spikenet Technology, LS2N, and Lagadic. Its aim was to develop vision-based localization and navigation techniques for autonomous landing on a runway (see Section 7.1.4).

# 9.2.3. ANR Contint Entracte

Participant: Julien Pettré.

#### no Inria Rennes 8013, duration: 42 months.

This project ended in April 2017. It was realized in collaboration with the Gepetto group at Laas, Toulouse, and the Mimetic group at Irisa and Inria Rennes Bretagne Atlantique. It addressed the problem of motion planning for anthropomorphic systems, and more generally, the problem of manipulation path planning. Entracte proposed to study in parallel both the mathematical foundations of artificial motion and the neurocognitive structures used by humans to quickly solve motion problems.

## 9.2.4. ANR JCJC Percolation

Participant: Julien Pettré.

no Inria Rennes 7991, duration: 42 months.

The ANR "Jeune Chercheur" Percolation project ended on June 2017. It aimed at designing perception-based crowd simulation algorithms. We developed agents able of perceiving their virtual environment through virtual sensors, and able to navigate in it, as well as to interact with the other agents.

## 9.2.5. ANR JCJC SenseFly

Participants: Thomas Bellavoir, Muhammad Usman, Paolo Robuffo Giordano.

no Irisa CNRS 50476, duration: 36 months.

The ANR "Jeune Chercheur" SenseFly project started in August 2015. Its goal is to advance the state-of-theart in multi-UAV in the design and implementation of fully decentralized and sensor-based group behaviors by only resorting to onboard sensing (mainly cameras and IMU) and local communication (e.g., Bluetooth communication, wireless networks). Topics such as individual flight control, formation control robust against sensor limitations (e.g., limited field of view, occlusions), distributed estimation of relative positions/bearings from local sensing, maintenance of architectural properties of a multi-UAV formation are studied in the project. Part of the platforms described in Section 6.8.5 has been purchased thanks to this grant.

## 9.2.6. ANR PLaTINUM

Participants: Eduardo Fernandez Moral, Vincent Drevelle, Patrick Rives.

no Inria Sophia 10204, duration: 42 months.

This project started in November 2015. It involves a consortium managed by Litis in Rouen with IGN Matis (Paris), Le2i (Le Creusot) and Lagadic group. It aims at proposing novel solutions to robust long-term mapping of urban environments.

# 9.2.7. BPI Romeo 2

Participants: Giovanni Claudio, Fabien Spindler, François Chaumette.

no Inria Rennes 7114, duration: 60 months.

This project ended in October 2017. It involved a large consortium managed by Softbank Robotics (ex Aldebaran Robotics) with Laas in Toulouse, Isir in Paris, Lirmm in Montpellier, Inria groups Lagadic, Bipop (Pierre-Brice Wieber), Flowers (Pierre-Yves Oudeyer), etc. It aimed at developing advanced control and perception functionalities to a humanoid robot. In this project, we developed visual manipulation and navigation tasks with Romeo and Pepper.

# 9.2.8. Equipex Robotex

Participants: Fabien Spindler, François Chaumette.

no Inria Rennes 6388, duration: 9 years.

Lagadic is one of the 15 French academic partners involved in the Equipex Robotex network that started in February 2011. It is devoted to get and manage significant equipment in the main robotics labs in France. In the scope of this project, we have got the humanoid robot Romeo (see Section 6.8.4).

# 9.3. European Initiatives

# 9.3.1. FP7 & H2020 Projects

9.3.1.1. FP7 Space RemoveDEBRIS

Participants: Eric Marchand, François Chaumette.

Instrument: Specific Targeted Research Project

Duration: October 2013 - September 2018

Coordinator: University of Surrey (United Kingdom)

Partners: Surrey Satellite Technology (United Kingdom), Airbus (Toulouse, France and Bremen, Germany), Isis (Delft, The Netherlands), CSEM (Neuchâtel, Switzerland), Stellenbosch University (South Africa).

Inria contact: François Chaumette

Abstract: The goal of this project is to validate model-based tracking algorithms on images acquired during an actual space debris removal mission [22], [47].

#### 9.3.1.2. H2020 ICT Comanoid

Participants: Giovanni Claudio, Souriya Trinh, Fabien Spindler, François Chaumette.

Title: Multi-contact Collaborative Humanoids in Aircraft Manufacturing

Programme: H2020

Duration: January 2015 - December 2018

Coordinator: CNRS (Lirmm)

Partners: Airbus Group (France), DLR (Germany), Università Degli Studi di Roma La Sapienza (Italy), CNRS (I3S)

Inria contact: Francois Chaumette

Comanoid investigates the deployment of robotic solutions in well-identified Airbus airliner assembly operations that are laborious or tedious for human workers and for which access is impossible for wheeled or rail-ported robotic platforms. As a solution to these constraints a humanoid robot is proposed to achieve the described tasks in real-use cases provided by Airbus Group. At a first glance, a humanoid robotic solution appears extremely risky, since the operations to be conducted are in highly constrained aircraft cavities with non-uniform (cargo) structures. Furthermore, these tight spaces are to be shared with human workers. Recent developments, however, in multi-contact planning and control suggest that this is a much more plausible solution than current alternatives such as a manipulator mounted on multi-legged base. Indeed, if humanoid robots can efficiently exploit their surroundings in order to support themselves during motion and manipulation, they can ensure balance and stability, move in non-gaited (acyclic) ways through narrow passages, and also increase operational forces by creating closed-kinematic chains. Bipedal robots are well suited to narrow environments specifically because they are able to perform manipulation using only small support areas. Moreover, the stability benefits of multi-legged robots that have larger support areas are largely lost when the manipulator must be brought close, or even beyond, the support borders. COMANOID aims at assessing clearly how far the state-of-the-art stands from such novel technologies. In particular the project focuses on implementing a real-world humanoid robotics solution using the best of research and innovation. The main challenge are to integrate current scientific and technological advances including multi-contact planning and control; advanced visual-haptic servoing; perception and localization; human-robot safety, and the operational efficiency of cobotics solutions in airliner manufacturing.

#### 9.3.1.3. H2020 ICT Romans

**Participants:** Firas Abi Farraj, Fabien Spindler, François Chaumette, Claudio Pacchierotti, Paolo Robuffo Giordano.

Title: Robotic Manipulation for Nuclear Sort and Segregation

Programme: H2020

Duration: May 2015 - April 2018

Coordinator: University of Birmingham

Partners: NLL (UK), CEA (France), Univ. Darmstadt (Germany)

CNRS contact: Paolo Robuffo Giordano

The RoMaNS (Robotic Manipulation for Nuclear Sort and Segregation) project will advance the state of the art in mixed autonomy for tele-manipulation, to solve a challenging and safety-critical "sort and segregate" industrial problem, driven by urgent market and societal needs. Cleaning up the past half century of nuclear waste, in the UK alone (mostly at the Sellafield site), represents the largest environmental remediation project in the whole of Europe. Most EU countries face related challenges. Nuclear waste must be "sorted and segregated", so that low-level waste is placed in low-level storage containers, rather than occupying extremely expensive and resource intensive high-level storage containers and facilities. Many older nuclear sites (>60 years in UK) contain large numbers of legacy storage containers, some of which have contents of mixed contamination levels, and sometimes unknown contents. Several million of these legacy waste containers must now be cut open, investigated, and their contents sorted. This can only be done remotely using robots, because

of the high levels of radioactive material. Current state-of-the-art practice in the industry, consists of simple tele-operation (e.g. by joystick or teach-pendant). Such an approach is not viable in the long-term, because it is prohibitively slow for processing the vast quantity of material required. The project aims at: 1) Develop novel hardware and software solutions for advanced bi-lateral master-slave tele-operation. 2) Develop advanced autonomy methods for highly adaptive automatic grasping and manipulation actions. 3) Combine autonomy and tele-operation methods using state-of-the-art understanding of mixed initiative planning, variable autonomy and shared control approaches. 4) Deliver a TRL 6 demonstration in an industrial plant-representative environment at the UK National Nuclear Lab Workington test facility.

## 9.3.2. Collaborations in European Programs, Except FP7 & H2020

#### 9.3.2.1. Interreg Adapt

Participants: Nicolas Le Borgne, Marie Babel.

Programme: Interreg VA France (Channel) England

Project acronym: Adapt

Project title: Assistive Devices for empowering disAbled People through robotic Technologies

Duration: 01/2017 - 06/2021

Coordinator: ESIGELEC/IRSEEM Rouen

Other partners: INSA Rennes - IRISA, LGCGM, IETR (France), Université de Picardie Jules Verne -MIS (France), Pôle Saint Hélier (France), CHU Rouen (France), Réseau Breizh PC (France), Ergovie (France), Pôle TES (France), University College of London - Aspire CREATE (UK), University of Kent (UK), East Kent Hospitals Univ NHS Found. Trust (UK), Health and Europe Centre (UK), Plymouth Hospitals NHS Trust (UK), Canterbury Christ Church University (UK), Kent Surrey Sussex Academic Health Science Network (UK), Cornwall Mobility Center (UK).

Abstract: This project aims to develop innovative assistive technologies in order to support the autonomy and to enhance the mobility of power wheelchair users with severe physical/cognitive disabilities. In particular, the objective is to design and evaluate a power wheelchair simulator as well as to design a multi-layer driving assistance system.

# 9.3.3. Collaborations with European Partners

#### 9.3.3.1. ANR Opmops

Participants: Florian Berton, Julien Pettré.

Programme: ANR

Project acronym: Opmops

Project title: Organized Pedestrian Movement in Public Spaces: Preparation and Crisis Management of Urban Parades and Demonstration Marches with High Conflict Potential

Duration: June 2017 - June 2020

Coordinator: Université de Haute Alsace (for France), Technische Universität Kaiserslautern (for Germany)

Other partners: Gendarmerie Nationale, Hochschule München, ONHYS S.A.S, Polizei Rheinland-Pfalz, Universität Koblenz-Landau, VdS GmbH

Abstract: This project is about parades of highly controversial groups or of political demonstration marches are considered as a major threat to urban security. Due to the movement of the urban parades and demonstration marches (in the following abbreviated by UPM) through large parts of cities and the resulting space and time dynamics, it is particularly difficult for forces of civil security (abbreviated in the following by FCS) to guarantee safety at these types of urban events without endangering one of the most important indicators of a free society. In this proposal, partners representing the FCS (police and industry) will cooperate with researchers from academic

institutions to develop a decision support tool which can help them both in the preparation phase and crisis management situations of UPMs. Specific technical issues which the French-German consortium will have to tackle include the following: Optimization methods to plan UPM routes, transportation to and from the UPM, location and personnel planning of FCS, control of UPMs using stationary and moving cameras, and simulation methods, including their visualization, with specific emphasis on social behavior.

#### 9.3.3.2. iProcess

Participants: Agniva Sengupta, Fabien Spindler, Eric Marchand, Alexandre Krupa, François Chaumette.

Project acronym: i-Process

Project title: Innovative and Flexible Food Processing Technology in Norway

Duration: January 2016 - December 2019

Coordinator: Sintef (Norway)

Other partners: Nofima, Univ. of Stavanger, NMBU, NTNU (Norway), DTU (Denmark), KU Leuven (Belgium), and about 10 Norwegian companies.

Abstract: This project is granted by the Norwegian Government. Its main objective is to develop novel concepts and methods for flexible and sustainable food processing in Norway. In the scope of this project, the Lagadic group is involved for visual tracking and visual servoing of generic and potentially deformable objects (see Section 7.1.2). Prof. Ekrem Misimi from Sintef spent a 4-month visit from May 2017 and a 1-week visit in November 2017. François Chaumette and Alexandre Krupa spent a short period at Sintef in Trondheim in February and June 2017 respectively.

# 9.4. International Initiatives

#### 9.4.1. Inria Associate Teams Not Involved in an Inria International Labs

9.4.1.1. SIMS

Title: Realistic and Efficient Simulation of Complex Systems

International Partners:

University of North Carolina at Chapel Hill (USA) - GAMMA Group - Ming C. Lin, Dinesh Manocha

University of Minnesota (USA) - Motion Lab - Stephen Guy

Brown University (USA) - VenLab - William Warren

Start year: 2012

See http://people.rennes.inria.fr/Julien.Pettre/EASIMS/easims.html

The general goal of SIMS is to make significant progress toward realistic and efficient simulation of highly complex systems, which raise combinatory explosive problems. This proposal is focused on human motion and interaction, and covers 3 active topics with wide application range:

- 1. Crowd simulation: virtual human interacting with other virtual humans,
- 2. Autonomous virtual humans interacting with their environment,
- 3. Physical simulation: real humans interacting with virtual environments.

SIMS is orthogonally structured by transversal questions: the evaluation of the level of realism reached by a simulation (which is a problem by itself in the considered topics), considering complex systems at various scales (micro, meso and macroscopic ones), and facing combinatory explosion of simulation algorithms.

#### 9.4.1.2. ISI4NAVE

Title: Innovative Sensors and adapted Interfaces for assistive NAVigation and pathology Evaluation International Partner (Institution - Laboratory - Researcher): University College London (United Kingdom) - Aspire CREATe - Tom Carlson Start year: 2016

See also: http://www.irisa.fr/lagadic/team/MarieBabel/ISI4NAVE/ISI4NAVE.html

The global ageing population, along with disability compensation constitutes major challenging societal and economic issues. In particular, achieving autonomy remains a fundamental need that contributes to the individual's wellness and well-being. In this context, innovative and smart technologies are designed to achieve independence while matching user's individual needs and desires.

Hence, designing a robotic assistive solution related to wheelchair navigation remains of major importance as soon as it compensates partial incapacities. This project then addresses the following two issues. First, the idea is to design an indoor / outdoor efficient obstacle avoidance system that respects the user intention, and does not alter user perception. This involves embedding innovative sensors to tackle the outdoor wheelchair navigation problem. The second objective is to take advantage of the proposed assistive tool to enhance the user Quality of Experience by means of biofeedback. Indeed, adapted interfaces should improve the understanding of people that suffer from cognitive and/or visual impairments.

The originality of the project is to continuously integrate medical validation as well as clinical trials during the scientific research work in order to match user needs and acceptation.

## 9.4.2. Participation in International Programs

#### 9.4.2.1. ACRV

The Lagadic group is one of the five external partners of the Australian Center for Robotic Vision (see http:// roboticvision.org). This center groups QUT in Brisbane, ANU in Canberra, Monash University and Adelaide University. In the scope of this project, Quentin Bateux received a grant to participate to the 2017 Robotic Vision Summer School in Kioloa (New South Wales) and spent a 1-week visit at QUT in March 2017.

# 9.5. International Research Visitors

# 9.5.1. Visits of International Scientists

- Prof. Denis Wolf, Associate Professor at Univ. Sao Paulo, Brazil, spent a sabbatical year in Sophia Antipolis from July 2016 to August 2017. He worked on semantic learning applied to intelligent vehicles.
- Prof. Dan Zelazo (Technion) and Prof. Antonio Bicchi (Univ. Pisa) spent a short visit in the group in Rennes in 2017.

9.5.1.1. Internships

- Giuseppe Sirignano (Univ. Salerno), from October 2017 to March 2018
- Mario Selvaggio (Univ. Naples), from November 2017 till end of December 2017

#### 9.5.2. Visits to International Teams

9.5.2.1. Research Stays Abroad

- Jason Chevrie spent a 3-month visit in Sarthak Misra's lab at the Surgical Robotics Laboratory (SRL) of University of Twente (Netherlands) where he performed robotic experiments in the scope of his Ph.D (see Section 7.3.2).
- François Chaumette was invited for a 1-week visit at Zhejiang University in November 2017.

# **10. Dissemination**

# **10.1. Promoting Scientific Activities**

# 10.1.1. Scientific Events Organisation

## 10.1.1.1. General Chair, Scientific Chair

- Marie Babel was the Scientific Chair of the workshop "Serious game et réalité virtuelle : de la conception à l'utilisation" organized in Inria Rennes on December 15th 2017.
- 10.1.1.2. Member of the Organizing Committees
  - Marie Babel and Patrick Rives were in the Organizing Committee of JNRR 2017 ("Journées Nationales de la Recherche en Robotique") held in Biarritz in November 2017
  - Alexandre Krupa co-organized with Christoph Hennersperger (Technical University of Munich) and Danail Stoyanov (University College London) the Workshop on "Medical Imaging Robotics" at IROS 2017, Vancouver, Canada (http://robotic-imaging.com)
  - Claudio Pacchierotti was the Publicity Chair of the 2017 IEEE World Haptics conference held in Münich, Germany. He also co-organized the workshop on "Wearable haptic systems: design, applications, and perspectives," at this conference.
  - Paolo Robuffo Giordano was Editor-in-Chief of the IEEE Int. Symposium on Multi-Robot and Multi-Agent Systems (MRS 2017)

#### 10.1.1.3. Chair of Conference Program Committees

• Julien Pettré was Program Chair for the 2017 ACM Motion in Games Conference, Barcelona, 8-10 November 2017

## 10.1.1.4. Member of the Conference Program Committees

- François Chaumette: ICRA 2018 (Associate Editor)
- Claudio Pacchierotti: WHC 2017 (Associate Editor)
- Julien Pettré: Eurographics 2017, ACM SCA 2017
- Patrick Rives: ICINCO 2017 (Program Committee Membership)
- Paolo Robuffo Giordano: IROS 2018 (Associate Editor), ICRA 2018 (Associate Editor), RSS 2018, ISRR 2017

#### 10.1.1.5. Reviewer

- Marie Babel: ICRA 2017 (1)
- François Chaumette: CDC 2017 (1)
- Vincent Drevelle: IROS 2017 (1)
- Alexandre Krupa: IROS 2017 (1), Surgetica 2017 (1), ICRA 2018 (1)
- Eric Marchand: IROS 2017 (3), ICRA 2018 (2)
- Claudio Pacchierotti: WHC 2017, ICRA 2017, IROS 2017, Haptics 2018, also Senior Reviewer for IEEE Robotics & Automation Society Young Reviewers Program (from 2016)
- Julien Pettré: ACM SIGGRAPH (2), ACM SIGGRAPH Asia 2017 (2), Eurographics 2018 (1), IEEE VR 2018 (1), ACM SCA 2017 (2)
- Patrick Rives: ICRA 2017 (3), CVPR 2017 (2), ICCV 2017, Icinco 2017 (3), IV 2017 (1), PPINV 2017 (3)
- Paolo Robuffo Giordano: ACC 2017 (1), HFR 2017(2), ICRA 2018(2), IROS 2017(2)
- Paolo Salaris: ACC 2017 (1), RSS 2017 (2), IROS 2017 (1)
- Fabien Spindler: IROS 2017 (1)

# 10.1.2. Journal

## 10.1.2.1. Member of the Editorial Boards

- François Chaumette: Editorial Board of the Int. Journal of Robotics Research, Funding Senior Editor of the IEEE Robotics and Automation Letters, Editorial Board of the Springer Tracts in Advanced Robotics, Board Member of the Springer Encyclopedia of Robotics, Senior Editor of the IEEE Trans. on Robotics (from December 2017).
- Alexandre Krupa and Eric Marchand are Associate Editors of the IEEE Robotics and Automation Letters.
- Julien Pettré: Editorial board of Collective Dynamics, Associate Editor for Computer Animation and Virtual Worlds

## 10.1.2.2. Reviewer - Reviewing Activities

- Marie Babel: IEEE RAL (1), IEEE THMS (1)
- François Chaumette: Journal of Intelligent & Robotics Systems (1), Int. Journal of Robotics Research (1)
- Alexandre Krupa: Int. Journal of Robotics Research (1)
- Eric Marchand: Int. Journal of Robotics Research (1)
- Claudio Pacchierotti: TOH (5), RAL (8), TBME (1), TMECH(1), IJMRCAS(1), Plos One (1)
- Julien Pettré: Plos One (1), Wiley's Computer Animation and Virtual Worlds (1), Current Directions in Psychological Science (1), Elsevier's Computer and Graphics (1), IEEE TVCG (1), Journal of Motor Behavior (1).
- Patrick Rives: IEEE-Robotics and Automation Letters (1)
- Paolo Robuffo Giordano: L-CSS(1), RAL (1), TRO (1), TSMC (1)
- Paolo Salaris: IEEE-Robotics and Automation Letters (1), IEEE Trans. on Control, Systems Technology (1), Int. Journal of Robotics Research (1), Int. Journal of Control, Automation and Systems IJCAS (1)

# 10.1.3. Invited Talks

- Marie Babel:
  - "Assistance and Service Robotics in a Human Environment Workshop" at IEEE IROS 2017 workshop about "Smart wheelchairs: the tomorrow's vehicles?".
  - "Navigation en milieu humain : comprendre pour mieux assister" at JNRR 2017
- François Chaumette: "Visual servoing with and without image processing", University of Zhejiang, December 2017.
- Vincent Drevelle: "Set-membership interval methods for observation robotics", ENSTA Bretagne, Brest, France, October 2017.
- Eric Marchand: "The dense approach for visual servoing", "Journées scientifiques" Inria, Sophia Antipolis, June 2017.
- Claudio Pacchierotti: "Wearable haptics for VR and AR". Max-Planck Institute for Intelligent Systems, Stuttgart, Germany, 2017.
- Julien Pettré:
  - "Microscopic crowd modeling and simulation: an holistic approach", Univ. Pompeu Fabra (Barcelona), May 2017
  - "Pedestrian Dynamics: Modeling, Validation and Calibration", Brown University ICERM workshop, August 2017
  - "Velocity-based algorithms for crowd simulation", workshop "Piétons et foules", Univ. Orsay, December 2017

- Paolo Robuffo Giordano:
  - "A New Look at Shared Control for Complex Robotics Systems". Ecole Normale Supérieure (ENS), Rennes, France, December 2017
  - "Blending Human Assistance and Local Autonomy for Advanced Telemanipulation". Humanoids 2017 Workshop on Towards Robust Grasping And Manipulation Skills For Humanoids, November 2017
  - "Recent Results on Shared Control for Human- assisted Telemanipulation". IROS 2017 Workshop on Human in-the-loop robotic manipulation: on the influence of the human role, September 2017
  - "Shared Control Architectures for Single and Multiple Robots". LS2N, Nantes, France, July 2017
  - "Graph-Theoretical Tools for Sensor-based Multi-Robot Applications". Robolog 2017 workshop, IRISA, France, June 2017
  - "Collective Control, State Estimation and Human Interaction for Quadrotors in Unstructured Environments". ICRA 2017 Workshop on Human Multi-Robot System Interaction, June 2017
  - "Recent results in Collective Control, State Estimation and Human Interaction for Quadrotors in Unstructured Environments". Aerial Robotics workshop, Paris, France, March 2017
- Fabien Spindler: "ViSP integration in ROS ecosystem", "Journée technique : retour d'expériences autour de ROS" organized by "Réseau métier des roboticiens et mécatroniciens", Cité Scientifique de l'Université de Lille, August 2017.

# 10.1.4. Leadership within the Scientific Community

- François Chaumette is a 2016-2019 elected member of the Administrative Committee of the IEEE Robotics and Automation Society. He is also a member of the Scientific Council of the CNRS INS2I and vice-president of the Tremplin-ERC ANR program.
- Claudio Pacchierotti is the Chair of the IEEE Technical Committee on Haptics.
- Philippe Martinet is vice-president of the "GdR Robotique". François Chaumette and Patrick Rives are members of its scientific council.

# 10.1.5. Scientific Expertise

- Marie Babel served as an expert for the International Mission of the French Research Ministry (MEIRIES).
- François Chaumette served as a Panel member for the ERC PE7 consolidator grants. He also served in the selection committee for a Professor position at UTC (Compiègne). He was also involved in the evaluation of a research proposal submitted to the Research Foundation of Flanders, and for an Associate Prof. position at the City University of Hong Kong.
- Claudio Pacchierotti was a member of the jury for the EuroHaptics Best PhD thesis award, for the Robotics Made in Italy Video Contest, and served for Hans Fischer Senior Fellowship at the Institute for Advanced Study of the Technical University of Munich (TUM-IAS).
- Paolo Robuffo Giordano was reviewer of ERC Starting Grants and Consolidator Grants for the EU, of ANR projects, and of NWO projects (Dutch funding agency).
- Patrick Rives served as member of the "Comité d'Evaluation" of the ANR Challenge "Malin" and of the ANR Challenge "Rose".

# 10.1.6. Research Administration

• François Chaumette serves as the president of the committee in charge of all the temporary recruitments ("Commission Personnel") at Inria Rennes-Bretagne Atlantique and Irisa. He is also a member of the Head team of Inria Rennes-Bretagne Atlantique.

- Eric Marchand served as secretary in the board of the "Association Française pour la Reconnaissance et l'Interprétation des Formes" (AFRIF). He is also in charge of the Irisa Ph.D. students in the committee in charge of all the temporary recruitments ("Commission Personnel") at Inria Rennes-Bretagne Atlantique and Irisa. He is in the board of the "Pôle Images et Réseaux" and in the board of "Ecole doctorale Matisse".
- Alexandre Krupa is a member of the CUMIR ("Commission des Utilisateurs des Moyens Informatiques pour la Recherche") of Inria Rennes-Bretagne Atlantique.
- Paolo Salaris is member of the "Comité de Suivi Doctoral (CSD)" of Inria Sophia Antipolis.

# **10.2. Teaching - Supervision - Juries**

# 10.2.1. Teaching

Marie Babel:

Master INSA2: "Robotics", 26 hours, M1, INSA Rennes

Master INSA1: "Architecture", 30 hours, L3, INSA Rennes

Master INSA2: "Computer science project", 30 hours, M1, INSA Rennes

Master INSA2: "Image analysis", 18 hours, M1, INSA Rennes

Master INSA1: "Remedial math courses", 50 hours, L3, INSA Rennes

François Chaumette:

Master ESIR3: "Visual servoing", 8 hours, M2, Ecole supérieure d'ingénieurs de Rennes

#### Vincent Drevelle:

Master ESIR2: "Real-time systems and RTOS", 24 hours, M1, Esir Rennes

Master GLA: "Terrain information systems", 14 hours, M2, Université de Rennes 1

Master Info: "Artificial intelligence", 24 hours, M1, Université de Rennes 1

Master Elec: "Connected and geolocalized embedded applications", 24 hours, M2, Université de Rennes 1

Licence Info: "Computer systems architecture", 24 hours, L1, Université de Rennes 1

Licence and Master Elec: "Electronics project", 38 hours, L3 and M1, Université de Rennes 1

Portail Info-Elec: "Discovering programming and electronics", 14 hours, L1, Université de Rennes 1

LabFab School Mobility: "Build your robot", 30 hours, DU

Licence Miage: "Computer programming", 78 hours, L3, Université de Rennes 1

Master Elec: "Instrumentation, localization, GPS", 4 hours, M2, Université de Rennes 1

Master Elec: "Data fusion", 20 hours, M2, Université de Rennes 1

#### Alexandre Krupa:

Master SIBM (Signals and Images in Biology and Medicine): "Medical robotics guided from images", 4.5 hours, M2, Université de Rennes 1, Brest and Angers

Master FIP TIC-Santé: "Ultrasound visual servoing", 6 hours, M2, Télécom Physique Strasbourg

Master INSA3: "Modeling and engineering for Biology and Health applications", 12 hours, M2, INSA Rennes

Master ESIR3: "Ultrasound visual servoing", 9 hours, M2, Esir Rennes

Eric Marchand:

Master Esir2: "Colorimetry", 24 hours, M1, Esir Rennes

Master Esir2: "Computer vision: geometry", 24 hours, M1, Esir Rennes

Master Esir3: "Special effects", 24 hours, M2, Esir Rennes

Master Esir3: "Computer vision: tracking and recognition", 24 hours, M2, Esir Rennes

Master MRI: "Computer vision", 24 hours, M2, Université de Rennes 1

Master MIA: "Augmented reality", 4 hours, M2, Université de Rennes 1

Julien Pettré:

INSA1: "Programmation Informatique", 40 hours, INSA Rennes

## 10.2.2. Supervision

- Ph.D.: Renato José Martins, "Robust navigation and control of an autonomous vehicle", defended in October 2017, supervised by Patrick Rives and Samuel Siqueira Bueno (CTI) [12]
- Ph.D.: Jason Chevrie, "Flexible needle steering using ultrasound visual servoing", defended in December 2017, supervised by Alexandre Krupa and Marie Babel [11]
- Ph.D. in progress: Fabrizio Schiano, "Collective control with onboard sensors for multiple quadrotor UAVs", to be defended in January 2018, supervised by Paolo Robuffo Giordano
- Ph.D. in progress: Pedro Patlan-Rosales, "A robotic control framework for quantitative ultrasound elastography", to be defended in January 2018, supervised by Alexandre Krupa
- Ph.D. in progress: Quentin Bateux, "Visual servoing from global descriptors", started in October 2014, supervised by Eric Marchand
- Ph.D. in progress: Lesley-Ann Duflot, "Visual servoing using shearlet transform", started in November 2014, supervised by Alexandre Krupa and Brahim Tamadazte (Minarob group at FEMTO-ST, Besançon)
- Ph.D. in progress: Firas Abi Farraj, "Shared Control Architectures for Visual Servoing Tasks", started in October 2015, supervised by Paolo Robuffo Giordano
- Ph.D. in progress: Salma Jiddi, "Analyses géométrique et photométrique pour des applications de réalité mixte", started in October 2015, supervised by Eric Marchand and Philippe Robert (Technicolor)
- Ph.D. in progress: Ide Flore Kenmogne Fokam, "Cooperative localization in multi-robot fleets using interval analysis", started in October 2015, supervised Vincent Drevelle and Eric Marchand
- Ph.D. in progress: Bryan Penin "Model predictive visual servoing for UAVS", started in October 2015, supervised by Paolo Robuffo Giordano and François Chaumette
- Ph.D. in progress: Guillaume Cortes, "Motion Capture", started in October 2015, supervised Eric Marchand and Anatole Lécuyer (Hybrid group).
- Ph.D. in progress: Muhammad Usman, "Robust Vision-Based Navigation for Quadrotor UAVs", started in October 2015, supervised by Paolo Robuffo Giordano
- Ph.D. in progress: Louise Devigne, "Contribution d'une aide technique robotique à l'évaluation de pathologies neurologiques : Application à la navigation d'un fauteuil roulant", started in November 2015, supervised by Marie Babel and Philippe Gallien (Pôle Saint Hélier)
- Ph.D. in progress: Quentin Delamare, "Algorithmes d'estimation et de commande pour des quadrirotors en interaction physique avec l'environnement", started in September 2016, supervised by Paolo Robuffo Giordano and Antonio Franchi (LAAS)
- Ph.D. in progress: Axel Lopez Gandia, "Data assimilation for synthetic vision-based crowd simulation algorithms", started in October 2016, supervised by Julien Pettré and François Chaumette

- Ph.D. in progress: Aline Baudry, "Contribution à la modélisation des fauteuils roulants pour l'amélioration de leur navigation en mode semi-autonome", started in October 2016, supervised by Marie Babel and Sylvain Guégan (Mechanical Engineering Dpt/LGCGM at Insa Rennes)
- Ph.D. in progress: Hadrien Gurnel, "Shared control of a biopsy needle from haptic and ultrasound visual feedback", started in October 2016, supervised by Alexandre Krupa, Maud Marchal (Hybrid group at Inria Rennes-Bretagne Atlantique and Irisa) and Laurent Launay (IRT b<>com)
- Ph.D. in progress: Dayana Hassan, "Plate-forme robotisée d'assistance aux personnes à mobilité réduite", started in November 2016, supervised by Paolo Salaris, Patrick Rives, and Frank Anjeaux (Axyn robotique)
- Ph.D. in progress: Agniva Sengupta, "Localization and characterization of objects with unknown shapes", started in January 2017, supervised by Alexandre Krupa and Eric Marchand
- Ph.D. in progress: Xavier De Tinguy de la Giroulière, "Conception de techniques d'interaction multisensorielles pour la manipulation dextre d'objets en réalité virtuelle", started in September 2017, supervised by Maud Marchal, Anatole Lécuyer (Hybrid group) and Claudio Pacchierotti
- Ph.D. in progress: Rahaf Rahal, "Mixed tactile-force feedback for safe and intuitive robotic teleoperation", started in October 2017, supervised by Paolo Robuffo Giordano and Claudio Pacchierotti
- Ph.D. in progress: Nicolas Le Borgne, "Contrôle partagé et navigation assistée d'un fauteuil roulant en extérieur", started in October 2017, supervised by Marie Babel
- Ph.D. in progress: Florian Berton, "Design of a virtual reality platform for studying immersion and behaviours in aggressive crowds", started in November 2017, supervised by Julien Pettré

## 10.2.3. External Ph.D. and HdR Juries

- François Chaumette: Abed Malti (HdR, reviewer, Tlecem University, Algeria), Guillaume Laurent (HdR, reviewer, Femto ST, Besançon), Sylvain Lanneau (Ph.D., member, LS2N, Nantes)
- Eric Marchand: Angélique Loesch (Ph.D., reviewer, Univ. Clermont Auvergne, CEA), Pierre Rolin (Ph.D., reviewer, Univ. de Nancy), Hristina Hristova (Ph.D., president, Univ. de Rennes 1)
- Paolo Robuffo Giordano: Alexandre Boeuf (Ph.D., reviewer, LAAS, Toulouse), Valerio Modugno (Ph.D., reviewer, University of Rome La Sapienza, Italy), Mohamed Sorour (Ph.D., reviewer, LIRMM, Montpellier)

# **10.3.** Popularization

- Due to the visibility of our experimental platforms, the team is often requested to present its research activities to students, researchers or industry. Our panel of demonstrations allows us to highlight recent results concerning the positioning of an ultrasound probe by visual servoing, grasping and dual arm manipulation by Romeo, vision-based shared control using our haptic device for object manipulation, the control of a fleet of quadrotors, vision-based detection and tracking for space navigation in a rendezvous context, the semi-autonomous navigation of a wheelchair, and augmented reality applications. Some of these demonstrations are available as videos on VispTeam YouTube channel (https://www.youtube.com/user/VispTeam/videos). This year there were among others, demonstrations organized for the new year's lab greetings, for the Inria industry journey focused on data and applications in Paris, for the "Semaine de l'Informatique Graphique et de la Réalité virtuelle", and for the Science and music journey.
- Marie Babel participated as a member in a panel discussion during the Digital Tech Conference organized in Rennes in December 2017.
- Julien Pettré wrote a vulgarization paper [68], explaining how crowd simulation is used in the field of visual effects for the movie and the video-game industries. The crowd simulation topic was presented during the "Pint of Science FR 2017" event, which organizes mini conferences animated by researchers in public places such as cafes and bars.

# 11. Bibliography

# Major publications by the team in recent years

- [1] F. CHAUMETTE, S. HUTCHINSON. *Visual servoing and visual tracking*, in "Handbook of Robotics", B. SICILIANO, O. KHATIB (editors), Springer, 2008, chap. 24, p. 563-583, http://hal.inria.fr/hal-00920414/en.
- [2] A. COMPORT, E. MARCHAND, M. PRESSIGOUT, F. CHAUMETTE.*Real-time markerless tracking for augmented reality: the virtual visual servoing framework*, in "IEEE Trans. on Visualization and Computer Graphics", July 2006, vol. 12, n<sup>O</sup> 4, p. 615–628, https://hal.inria.fr/inria-00161250.
- [3] A. DAME, E. MARCHAND.Second order optimization of mutual information for real-time image registration, in "IEEE Trans. on Image Processing", 2012, vol. 21, n<sup>o</sup> 9, p. 4190-4203, http://hal.inria.fr/hal-00750528/en.
- [4] A. DIOSI, S. SEGVIC, A. REMAZEILLES, F. CHAUMETTE. Experimental Evaluation of Autonomous Driving Based on Visual Memory and Image Based Visual Servoing, in "IEEE Trans. on Intelligent Transportation Systems", September 2011, vol. 12, n<sup>o</sup> 3, p. 870–883, http://hal.inria.fr/hal-00639680/en.
- [5] E. MARCHAND, F. SPINDLER, F. CHAUMETTE. ViSP for visual servoing: a generic software platform with a wide class of robot control skills, in "IEEE Robotics and Automation Magazine", December 2005, vol. 12, n<sup>o</sup> 4, p. 40-52, https://hal.inria.fr/inria-00351899.
- [6] R. MEBARKI, A. KRUPA, F. CHAUMETTE.2D ultrasound probe complete guidance by visual servoing using image moments, in "IEEE Trans. on Robotics", April 2010, vol. 26, n<sup>o</sup> 2, p. 296-306, https://hal.inria.fr/inria-00544791.
- [7] M. MEILLAND, A. COMPORT, P. RIVES. Dense omnidirectional RGB-D mapping of large scale outdoor environments for real-time localisation and autonomous navigation, in "Journal of Field Robotics, Special Issue on Ground Robots Operating in dynamic, unstructured and large-scale outdoor environments", June 2015, vol. 32, n<sup>o</sup> 4, p. 474-503, http://hal.inria.fr/hal-01010429.
- [8] C. NADEAU, A. KRUPA.Intensity-based ultrasound visual servoing: modeling and validation with 2D and 3D probes, in "IEEE. Trans. on Robotics", August 2013, vol. 29, n<sup>o</sup> 4, p. 1003-1015 [DOI: 10.1109/TRO.2013.2256690], http://hal.inria.fr/hal-00854100.
- [9] R. SPICA, P. ROBUFFO GIORDANO, F. CHAUMETTE. Active Structure from Motion: Application to Point, Sphere and Cylinder, in "IEEE Trans. on Robotics", December 2015, vol. 30, n<sup>o</sup> 6, p. 1499-1513, http://hal. inria.fr/hal-01010429.
- [10] D. ZELAZO, A. FRANCHI, H. H. BÜLTHOFF, P. ROBUFFO GIORDANO. Decentralized rigidity maintenance control with range measurements for multi-robot systems, in "The Int. Journal of Robotics Research", January 2015, vol. 34, n<sup>o</sup> 1, p. 105-128, http://hal.inria.fr/hal-01076423.

# **Publications of the year**

# **Doctoral Dissertations and Habilitation Theses**

[11] J. CHEVRIE.*Flexible Needle Steering using Ultrasound Visual Servoing*, Université de Rennes 1, December 2017, https://tel.archives-ouvertes.fr/tel-01663761.

[12] R. MARTINS.Direct visual odometry and dense large-scale environment mapping from panoramic RGB-D images, Université de recherche Paris Sciences et Lettres ; Mines Paristech, October 2017, https://tel.archivesouvertes.fr/tel-01657421.

#### **Articles in International Peer-Reviewed Journal**

- [13] D. J. AGRAVANTE, G. CLAUDIO, F. SPINDLER, F. CHAUMETTE. Visual servoing in an optimization framework for the whole-body control of humanoid robots, in "IEEE Robotics and Automation Letters", April 2017, vol. 2, n<sup>o</sup> 2, p. 608-615, https://hal.inria.fr/hal-01421734.
- [14] G. ANTONELLI, E. CATALDI, F. ARRICHIELLO, P. ROBUFFO GIORDANO, S. CHIAVERINI, A. FRANCHI.Adaptive Trajectory Tracking for Quadrotor MAVs in Presence of Parameter Uncertainties and External Disturbances, in "IEEE Transactions on Control Systems Technology", January 2018, vol. 26, n<sup>O</sup> 1, p. 248-254 [DOI: 10.1109/TCST.2017.2650679], https://hal.inria.fr/hal-01483653.
- [15] Q. BATEUX, E. MARCHAND. *Histograms-based Visual Servoing*, in "IEEE Robotics and Automation Letters", January 2017, vol. 2, n<sup>o</sup> 1, p. 80-87, https://hal.inria.fr/hal-01265560.
- [16] S. BRIOT, F. CHAUMETTE, P. MARTINET. Revisiting the determination of the singularity cases in the visual servoing of image points through the concept of hidden robot, in "IEEE Transactions on Robotics", January 2017, vol. 33, n<sup>o</sup> 2, https://hal.archives-ouvertes.fr/hal-01399774.
- [17] S. BRIOT, P. MARTINET, F. CHAUMETTE. Determining the Singularities for the Observation of Three Image Lines, in "IEEE Robotics and Automation Letters", April 2017, vol. 2, n<sup>o</sup> 2, p. 412-419, https://hal.archivesouvertes.fr/hal-01398925.
- [18] J. BRUNEAU, J. PETTRÉ. EACS: Effective Avoidance Combination Strategy, in "Computer Graphics Forum", December 2017, vol. 36, n<sup>o</sup> 8, p. 108-122 [DOI: 10.1111/CGF.13066], https://hal.inria.fr/hal-01392248.
- [19] P. CHATELAIN, A. KRUPA, N. NAVAB. *Confidence-Driven Control of an Ultrasound Probe*, in "IEEE Transactions on Robotics", December 2017, vol. 33, n<sup>o</sup> 6, p. 1410-1424, https://hal.inria.fr/hal-01551431.
- [20] F. CHINELLO, C. PACCHIEROTTI, M. MALVEZZI, D. PRATTICHIZZO. A Three Revolute-Revolute-Spherical wearable fingertip cutaneous device for stiffness rendering, in "IEEE Transactions on Haptics (ToH)", 2017 [DOI: 10.1109/TOH.2017.2755015], https://hal.inria.fr/hal-01616509.
- [21] T. B. DUTRA, R. MARQUES, J. B. B. CAVALCANTE-NETO, C. A. VIDAL, J. PETTRÉ.Gradient-based steering for vision-based crowd simulation algorithms, in "Computer Graphics Forum", May 2017, vol. 36, n<sup>0</sup> 2, p. 337 - 348 [DOI: 10.1111/CGF.13130], https://hal.inria.fr/hal-01653590.
- [22] J. L. FORSHAW, G. S. AGLIETTI, T. SALMON, I. RETAT, M. ROE, C. BURGESS, T. CHABOT, A. PISSELOUP, A. PHIPPS, C. BERNAL, F. CHAUMETTE, A. POLLINI, W. H. STEYN. *Final payload test results* for the RemoveDebris active debris removal mission, in "Acta Astronautica", September 2017, vol. 138, p. 326 - 342 [DOI: 10.1016/J.ACTAASTRO.2017.06.003], https://hal.inria.fr/hal-01546097.
- [23] A. FRANCHI, P. ROBUFFO GIORDANO. Online Leader Selection for Improved Collective Tracking and Formation Maintenance, in "IEEE transactions on control of network systems", 2017 [DOI: 10.1109/TCNS.2016.2567222], https://hal.archives-ouvertes.fr/hal-01315463.

- [24] M. MAISTO, C. PACCHIEROTTI, F. CHINELLO, G. SALVIETTI, A. DE LUCA, D. PRATTICHIZZO. Evaluation of wearable haptic systems for the fingers in Augmented Reality applications, in "IEEE Transactions on Haptics (ToH)", 2017 [DOI: 10.1109/TOH.2017.2691328], https://hal.inria.fr/hal-01616505.
- [25] L. MELI, C. PACCHIEROTTI, D. PRATTICHIZZO. Experimental evaluation of magnified haptic feedback for robot-assisted needle insertion and palpation, in "The International Journal of Medical Robotics and Computer Assisted Surgery", December 2017, vol. 13, n<sup>o</sup> 4, p. 1-14 [DOI: 10.1002/RCS.1809], https://hal.inria.fr/hal-01482249.
- [26] T. NESTMEYER, P. ROBUFFO GIORDANO, H. H. BÜLTHOFF, A. FRANCHI. Decentralized Simultaneous Multi-target Exploration using a Connected Network of Multiple Robots, in "Autonomous Robots", April 2017, vol. 41, nº 4, p. 989-1011 [DOI: 10.1007/s10514-016-9578-9], https://hal.inria.fr/hal-01332937.
- [27] A.-H. OLIVIER, J. BRUNEAU, R. KULPA, J. PETTRÉ. Walking with virtual people: Evaluation of locomotion interfaces in dynamic environments, in "IEEE Transactions on Visualization and Computer Graphics", 2017 [DOI: 10.1109/TVCG.2017.2714665], https://hal.inria.fr/hal-01557761.
- [28] C. PACCHIEROTTI, F. ONGARO, F. VAN DEN BRINK, C. YOON, D. PRATTICHIZZO, D. H. GRA-CIAS, S. MISRA. Steering and Control of Miniaturized Untethered Soft Magnetic Grippers With Haptic Assistance, in "IEEE Transactions on Automation Science and Engineering", 2017, p. 1 - 17 [DOI: 10.1109/TASE.2016.2635106], https://hal.inria.fr/hal-01482255.
- [29] C. PACCHIEROTTI, S. SINCLAIR, M. SOLAZZI, A. FRISOLI, V. HAYWARD, D. PRATTICHIZZO. Wearable Haptic Systems for the Fingertip and the Hand: Taxonomy, Review, and Perspectives, in "IEEE Transactions on Haptics (ToH)", 2017 [DOI: 10.1109/TOH.2017.2689006], https://hal.inria.fr/hal-01616501.
- [30] P. PAPADAKIS, P. RIVES. Binding human spatial interactions with mapping for enhanced mobility in dynamic environments, in "Autonomous Robots", May 2017, vol. 41, n<sup>o</sup> 5, p. 1047-1059 [DOI: 10.1007/s10514-016-9581-1], https://hal.inria.fr/hal-01342255.
- [31] A. PEREZ-YUS, E. FERNANDEZ-MORAL, G. LOPEZ-NICOLAS, J. J. GUERRERO, P. RIVES. Extrinsic Calibration of Multiple RGB-D Cameras From Line Observations, in "IEEE Robotics and Automation Letters", 2018, vol. 3, n<sup>o</sup> 1, p. 273 - 280 [DOI : 10.1109/LRA.2017.2739104], https://hal.inria.fr/hal-01581523.
- [32] Z. REN, P. CHARALAMBOUS, J. BRUNEAU, Q. PENG, J. PETTRÉ.Group Modeling: A Unified Velocity-Based Approach, in "Computer Graphics Forum", December 2017, vol. 36, n<sup>o</sup> 8, p. 45-56 [DOI: 10.1111/CGF.12993], https://hal.inria.fr/hal-01372766.
- [33] L. ROYER, A. KRUPA, G. DARDENNE, A. LE BRAS, E. MARCHAND, M. MARCHAL.Realtime Target Tracking of Soft Tissues in 3D Ultrasound Images Based on Robust Visual Information and Mechanical Simulation, in "Medical Image Analysis", January 2017, vol. 35, p. 582 - 598 [DOI: 10.1016/J.MEDIA.2016.09.004], https://hal.inria.fr/hal-01374589.
- [34] P. SALARIS, N. ABE, J.-P. LAUMOND. Robot Choreography: The Use of the Kinetography Laban System to Notate Robot Action and Motion, in "IEEE Robotics and Automation Magazine", September 2017, vol. 24, n<sup>o</sup> 3, p. 30 - 40 [DOI: 10.1109/MRA.2016.2636361], https://hal.inria.fr/hal-01623920.
- [35] R. SPICA, P. ROBUFFO GIORDANO, F. CHAUMETTE. Coupling Active Depth Estimation and Visual Servoing via a Large Projection Operator, in "International Journal of Robotics Research", 2017, vol. 36, n<sup>o</sup> 11, p. 1177-1194, https://hal.inria.fr/hal-01572366.
- [36] C. VASSALLO, A.-H. OLIVIER, P. SOUÈRES, A. CRÉTUAL, O. STASSE, J. PETTRÉ. How do walkers avoid a mobile robot crossing their way?, in "Gait and Posture", January 2017, vol. 51, p. 97-103, https://arxiv.org/ abs/1609.07955 [DOI: 10.1016/J.GAITPOST.2016.09.022], https://hal.archives-ouvertes.fr/hal-01371202.

### **International Conferences with Proceedings**

- [37] F. ABI-FARRAJ, T. OSA, N. PEDEMONTE, J. PETERS, G. NEUMANN, P. ROBUFFO GIORDANO. A Learningbased Shared Control Architecture for Interactive Task Execution, in "IEEE Int. Conf. on Robotics and Automation, ICRA'17", Singapore, Singapore, May 2017, https://hal.inria.fr/hal-01482137.
- [38] D. J. AGRAVANTE, F. CHAUMETTE. Active vision for pose estimation applied to singularity avoidance in visual servoing, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'17", Vancouver, Canada, September 2017, p. 2947-2952, https://hal.inria.fr/hal-01589882.
- [39] N. BATTILANI, R. SPICA, P. ROBUFFO GIORDANO, C. SECCHI.An Assisted Bilateral Control Strategy for 3D Pose Estimation of Visual Features, in "IROS 2017 - IEEE/RSJ International Conference on Intelligent Robots and Systems", Vancouver, Canada, September 2017, p. 2580-2586, https://hal.inria.fr/hal-01572356.
- [40] J. BIMBO, C. PACCHIEROTTI, M. AGGRAVI, N. TSAGARAKIS, D. PRATTICHIZZO. *Teleoperation in cluttered environments using wearable haptic feedback*, in "IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS'17", Vancouver, Canada, September 2017, p. 3401-3408, https://hal.inria.fr/hal-01572358.
- [41] N. CAZY, P.-B. WIEBER, P. ROBUFFO GIORDANO, F. CHAUMETTE. Visual Servoing Using Model Predictive Control to Assist Multiple Trajectory Tracking, in "IEEE Int. Conf. on Robotics and Automation, ICRA'17", Singapore, Singapore, May 2017, https://hal.inria.fr/hal-01482878.
- [42] G. CORTES, E. MARCHAND, J. ARDOUIN, A. LÉCUYER. An Optical Tracking System based on Hybrid Stereo/Single-View Registration and Controlled Cameras, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'17", Vancouver, Canada, September 2017, p. 6185-6190, https://hal.inria.fr/hal-01562327.
- [43] G. CORTES, E. MARCHAND, J. ARDOUIN, A. LÉCUYER. Increasing Optical Tracking Workspace of VR Applications using Controlled Cameras, in "IEEE Symposium on 3D User Interfaces, 3DUI 2017", Los Angeles, United States, March 2017, p. 22-25 [DOI: 10.1109/3DUI.2017.7893313], https://hal.inria.fr/ hal-01446343.
- [44] L. DEVIGNE, M. BABEL, F. NOUVIALE, V. K. NARAYANAN, F. PASTEAU, P. GALLIEN. Design of an immersive simulator for assisted power wheelchair driving, in "IEEE Int. Conf. on Rehabilitation Robotics, ICORR'17", London, United Kingdom, July 2017, https://hal.inria.fr/hal-01523385.
- [45] C. EZEH, P. TRAUTMAN, L. DEVIGNE, V. BUREAU, M. BABEL, T. CARLSON. Probabilistic vs Linear Blending Approaches to Shared Control for Wheelchair Driving, in "IEEE Int. Conf. on Rehabilitation Robotics, ICORR'17", London, United Kingdom, July 2017, https://hal.inria.fr/hal-01523412.

- [46] E. FERNANDEZ-MORAL, R. MARTINS, D. WOLF, P. RIVES. A new metric for evaluating semantic segmentation: leveraging global and contour accuracy, in "Workshop on Planning, Perception and Navigation for Intelligent Vehicles, PPNIV17", Vancouver, Canada, September 2017, https://hal.inria.fr/hal-01581525.
- [47] J. L. FORSHAW, G. S. AGLIETTI, T. SALMON, I. RETAT, C. BURGESS, T. CHABOT, A. PISSELOUP, A. PHIPPS, C. BERNAL, A. POLLINI, W. H. STEYN, F. CHAUMETTE. *The RemoveDebris ADR Mission: Preparing for an International Space Station Launch*, in "7th European Conference on Space Debris", Darmstadt, Germany, April 2017, https://hal.inria.fr/hal-01576084.
- [48] A. M. GHALAMZAN, F. ABI-FARRAJ, P. ROBUFFO GIORDANO, R. STOLKIN. Human-in-the-loop optimisation: mixed initiative grasping for optimally facilitating post-grasp manipulative actions, in "IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS'17", Vancouver, Canada, September 2017, p. 3386-3393, https://hal.inria.fr/hal-01572347.
- [49] I.-F. KENMOGNE, V. DREVELLE, E. MARCHAND.*Image-based UAV localization using Interval Methods*, in "IROS 2017 - IEEE/RSJ International Conference on Intelligent Robots and Systems", Vancouver, Canada, September 2017, p. 5285-5291, https://hal.inria.fr/hal-01572369.
- [50] I.-F. KENMOGNE, V. DREVELLE, E. MARCHAND. Vision based Pose domain characterization of an Unmanned Aerial Vehicle using Interval Analysis, in "10th Workshop on Interval Methods", Manchester, United Kingdom, June 2017, https://hal.inria.fr/hal-01540212.
- [51] E. MARCHAND, F. CHAUMETTE. Visual Servoing through mirror reflection, in "ICRA'17 IEEE International Conference on Robotics and Automation", Singapore, Singapore, May 2017, https://hal.inria.fr/hal-01445484.
- [52] E. MARCHAND, B. FASQUELLE. Visual Servoing from lines using a planar catadioptric system, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'17", Vancouver, Canada, September 2017, p. 2935-2940, https://hal.inria.fr/hal-01561757.
- [53] R. MARTINS, E. FERNANDEZ-MORAL, P. RIVES. An Efficient Rotation and Translation Decoupled Initialization from Large Field of View Depth Images, in "IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS'17", Vancouver, Canada, September 2017, p. 5750-5755, https://hal.inria.fr/hal-01581524.
- [54] L. MEERHOFF, J. PETTRÉ, R. KULPA, S. LYNCH, A. CRÉTUAL, A.-H. OLIVIER. Simultaneous and sequential affordances of collision avoidance between multiple pedestrians, in "ICPA 2017 - 19th biannual International Conference on Perception and Action", Séoul, South Korea, July 2017, https://hal.inria.fr/hal-01646507.
- [55] A.-H. OLIVIER, J. BRUNEAU, R. KULPA, J. PETTRÉ. Walking with virtual people: Evaluation of locomotion interfaces in dynamic environments, in "IEEE VR 2017 - IEEE Virtual Reality", Los Angeles, United States, IEEE, March 2017, vol. PP, n<sup>o</sup> 99, p. 1-13 [DOI: 10.1109/TVCG.2017.2714665], https://hal.inria.fr/hal-01661211.
- [56] P. A. PATLAN-ROSALES, A. KRUPA. A robotic control framework for 3-D quantitative ultrasound elastography, in "IEEE Int. Conf. on Robotics and Automation, ICRA'17", Singapore, Singapore, May 2017, https:// hal.inria.fr/hal-01481975.

- [57] P. A. PATLAN-ROSALES, A. KRUPA.Strain estimation of moving tissue based on automatic motion compensation by ultrasound visual servoing, in "IROS 2017 - IEEE/RSJ International Conference on Intelligent Robots and Systems", Vancouver, Canada, September 2017, p. 2941-2946, https://hal.inria.fr/hal-01572364.
- [58] N. PEDEMONTE, F. ABI-FARRAJ, P. ROBUFFO GIORDANO. Visual-Based Shared Control for Remote Telemanipulation with Integral Haptic Feedback, in "IEEE Int. Conf. on Robotics and Automation, ICRA'17", Singapore, Singapore, May 2017, https://hal.inria.fr/hal-01482129.
- [59] B. PENIN, R. SPICA, P. ROBUFFO GIORDANO, F. CHAUMETTE. Vision-Based Minimum-Time Trajectory Generation for a Quadrotor UAV, in "IROS 2017 - IEEE/RSJ International Conference on Intelligent Robots and Systems", Vancouver, Canada, September 2017, p. 6199-6206, https://hal.inria.fr/hal-01572362.
- [60] S. RAJ BISTA, P. ROBUFFO GIORDANO, F. CHAUMETTE. Combining Line Segments and Points for Appearance- based Indoor Navigation by Image Based Visual Servoing, in "IROS 2017 - IEEE/RSJ International Conference on Intelligent Robots and Systems", Vancouver, Canada, September 2017, p. 2960-2967, https://hal.inria.fr/hal-01572353.
- [61] P. SALARIS, R. SPICA, P. ROBUFFO GIORDANO, P. RIVES. Online Optimal Active Sensing Control, in "International Conference on Robotics and Automation (ICRA)", Singapore, Singapore, May 2017, https:// hal.inria.fr/hal-01472608.
- [62] F. SCHIANO, P. ROBUFFO GIORDANO. Bearing rigidity maintenance for formations of quadrotor UAVs, in "ICRA 2017 - IEEE International Conference on Robotics and Automation", Singapore, Singapore, May 2017, p. 1467 - 1474 [DOI: 10.1109/ICRA.2017.7989175], https://hal.inria.fr/hal-01482422.

### **Conferences without Proceedings**

- [63] Q. BATEUX, E. MARCHAND, J. LEITNER, F. CHAUMETTE, P. CORKE. Visual Servoing from Deep Neural Networks, in "RSS 2017 - Robotics : Science and Systems, Workshop New Frontiers for Deep Learning in Robotics", Boston, United States, July 2017, p. 1-6, https://hal.inria.fr/hal-01589887.
- [64] S. BRIOT, F. CHAUMETTE, P. MARTINET. Revisiting the determination of the singularity cases in the visual servoing of image points through the concept of hidden robot, in "ICRA 2017 - IEEE International Conference on Robotics and Automation", Singapour, Singapore, ICRA 2017, May 2017, n<sup>o</sup> 99 [DOI: 10.1109/TRO.2016.2637912], https://hal.archives-ouvertes.fr/hal-01435810.
- [65] S. BRIOT, P. MARTINET, F. CHAUMETTE. Singularity Cases in the Visual Servoing of Three Image Lines, in "2017 IEEE International Conference on Robotics and Automation (ICRA 2017)", Singapour, Singapore, Proceedings of 2017 IEEE International Conference on Robotics and Automation (ICRA 2017), May 2017, https://hal.archives-ouvertes.fr/hal-01435811.
- [66] S. LYNCH, J. PETTRÉ, J. BRUNEAU, R. KULPA, A. CRÉTUAL, A.-H. OLIVIER. Effect of mutual gaze on collision avoidance behavior during human walking, in "AFRV", Rennes, France, October 2017, https://hal. archives-ouvertes.fr/hal-01645801.

### Scientific Books (or Scientific Book chapters)

[67] A. BERA, D. WOLINSKI, J. PETTRÉ, D. MANOCHA. Realtime Pedestrian Tracking and Prediction in Dense Crowds, in "Group and Crowd Behavior for Computer Vision", Elsevier, April 2017, https://hal.inria.fr/hal-01653602.

### **Scientific Popularization**

[68] L. HOYET, J. PETTRÉ.La simulation de foule au service du divertissement, in "Interstices", April 2017, https:// hal.inria.fr/hal-01533687.

### **Other Publications**

- [69] G. CORTES, F. ARGELAGUET SANZ, E. MARCHAND, A. LÉCUYER. Toward Application-Driven VR Systems: Analysis of Head and Hand 3D Motions in a Specific CAVE-based Application, March 2017, IEEE Symposium on 3D User Interfaces, 3DUI 2017, poster session, Poster, https://hal.inria.fr/hal-01482150.
- [70] S. JIDDI, P. ROBERT, E. MARCHAND.Illumination Estimation using Cast Shadows for Realistic Augmented Reality Applications, October 2017, IEEE Int. Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), Poster, https://hal.inria.fr/hal-01668701.
- [71] S. LYNCH, R. KULPA, L. MEERHOFF, J. PETTRÉ, A. CRÉTUAL, A.-H. OLIVIER. Effect of local limb cues in the prediction of global motion during collision avoidance, October 2017, ACAPS, Poster, https://hal.archivesouvertes.fr/hal-01645924.
- [72] P. A. PATLAN-ROSALES, A. KRUPA. *A general framework for automatic robotic palpation*, September 2017, Workshop on Medical Imaging Robotics, IROS'17, https://hal.inria.fr/hal-01611325.
- [73] S. TONNEAU, A. D. PRETE, J. PETTRÉ, N. MANSARD.2PAC: Two Point Attractors for Center of Mass Trajectories in Multi Contact Scenarios, September 2017, working paper or preprint, https://hal.archivesouvertes.fr/hal-01609055.

# **Project-Team LINKMEDIA**

## Creating and exploiting explicit links between multimedia fragments

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: CNRS Institut national des sciences appliquées de Rennes Université Rennes 1

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Vision, perception and multimedia interpretation

### **Table of contents**

1.	Personnel	621
2.	Overall Objectives	622
	2.1. Context	622
	2.2. Scientific objectives	622
3.	Research Program	623
	3.1. Scientific background	623
	3.2. Workplan	623
	3.2.1. Unsupervised motif discovery	623
	3.2.2. Description and structuring	624
	3.2.3. Linking and collection data model	624
4.	Application Domains	625
	4.1. Asset management in the entertainement business	625
	4.2. Multimedia Internet	625
	4.3. Multiscreen TV	625
	4.4. E-learning	625
5.	Highlights of the Year	625
6.	New Software and Platforms	626
7.	New Results	626
	7.1. Multimedia indexing, Motif and knowledge discovery	626
	7.1.1. Towards engineering a web-scale multimedia service: a case study using SPARK	626
	7.1.2. On competitiveness of nearest-neighbor based music classification: a methodolog	gical
	critique	626
	7.1.3. Unsupervised part learning for visual recognition	627
	7.1.4. Automatic discovery of discriminative parts as a quadratic assignment problem	627
	7.1.5. Learning DTW-preserving shapelets	627
	7.1.6. Tag propagation approaches within speaking face graphs for multimodal person disco	very
		628
	7.2. Multimedia content description and structuring	628
	7.2.1. The vulnerability of learning to adversarial perturbation increases with intrinsic dir	nen-
	sionality	628
	7.2.2. Efficient temporal kernels between feature sets for time series classification	628
	7.2.3. Tampering detection and localization in images from social networks	629
	7.2.4. Identity documents classification as an image classification problem	629
	7.2.5. Sentiment analysis	630
	7.3. Content-based information retrieval	630
	7.3.1. Efficient diffusion on region manifolds: recovering small objects with compact (	CNN
	representations	630
	7.3.2. Panorama to panorama matching for location recognition	630
	7.3.3. Memory vectors for similarity search in high-dimensional spaces	630
	7.3.4. Exploiting multimodality in video hyperlinking to improve target diversity	631
	7.3.5. Generative adversarial networks for multimodal representation learning in video hy	/per-
	linking	631
	7.4. Linking, navigation and analytics	631
	7.4.1. Providing real-time insight during political debates in a second screen application	631
	7.4.2. Information extraction in clinical documents	632
	7.4.3. Semi-supervision for information extraction	632
	7.4.4. Linking multimedia content for efficient news browsing via explorable news graphs	632
	7.4.5. Multimodal detection of fake news	633
	7.5. Miscellaneous	633

	7.5.1. One-step time-dependent future video frame prediction with a convolutional encod	er-
	decoder neural network	633
	7.5.2. About zero bit watermarking error exponents	633
8.	Bilateral Contracts and Grants with Industry	633
9.	Partnerships and Cooperations	634
	9.1. Regional Initiatives	634
	9.1.1. CominLabs Project Linking Media in Acceptable Hypergraphs (LIMAH)	634
	9.1.2. CominLabs Project BigCLIN	634
	9.2. National Initiatives	634
	9.2.1. ANR Project IDFRAud	634
	9.2.2. FUI 19 NexGenTV	635
	9.2.3. Inria Project Lab Knowledge-driven data and content collaborative analytics (iCODA)	635
	9.3. European Initiatives	635
	9.3.1. CHIST ERA ID_IOT	635
	9.3.2. Collaborations with Major European Organizations	636
	9.4. International Initiatives	636
	9.4.1. Inria International Partners	636
	9.4.2. Participation in Other International Programs	636
	9.5. International Research Visitors	636
	9.5.1. Visits of International Scientists	636
	9.5.2. Visits to International Teams	636
10.	Dissemination	637
	10.1. Promoting Scientific Activities	637
	10.1.1. Scientific Events Organisation	637
	10.1.1.1. General Chair, Scientific Chair	637
	10.1.1.2. Member of the Organizing Committees	637
	10.1.2. Scientific Events Selection	637
	10.1.2.1. Chair of Conference Program Committees	637
	10.1.2.2. Member of Conference Program Committees	637
	10.1.2.3. Reviewer	638
	10.1.3.  Journal	638
	10.1.3.1. Member of the Editorial Boards	038
	10.1.4. Invite d Talles	638
	10.1.4. Invited faits	639
	10.1.5. Leadership within the Scientific Community	639
	10.1.7. Descent Administration	639
	10.1.7. Research Administration	039
	10.2. Teaching - Supervision - Juries	040 640
	10.2.1. Teaching	04U
	10.2.2. Supervision	041 641
	10.2. Dopularization	642
11	IU.S. FUPUIAIIZAUUII Dibliography	042 642
11.	ыонодгарну	042

### **Project-Team LINKMEDIA**

Creation of the Project-Team: 2014 July 01

### **Keywords:**

### **Computer Science and Digital Science:**

- A3.3.2. Data mining
- A3.3.3. Big data analysis
- A5.3.3. Pattern recognition
- A5.4.1. Object recognition
- A5.4.3. Content retrieval
- A5.7. Audio modeling and processing
- A5.8. Natural language processing
- A9.2. Machine learning
- A9.4. Natural language processing

### **Other Research Topics and Application Domains:**

B9. - Society and Knowledge

### **1. Personnel**

### **Research Scientists**

Laurent Amsaleg [CNRS, Researcher, HDR] Yannis Avrithis [Inria, Advanced Research Position] Vincent Claveau [CNRS, Researcher] Teddy Furon [Inria, Researcher] Guillaume Gravier [Team leader, CNRS, Senior Researcher, HDR] Miaojing Shi [Inria, Starting Research Position, from Dec 2017]

### **Faculty Members**

Ewa Kijak [Univ de Rennes I, Associate Professor] Simon Malinowski [Univ de Rennes I, Associate Professor] Christian Raymond [INSA Rennes, Associate Professor] Pascale Sébillot [INSA Rennes, Professor, HDR]

### **Technical Staff**

Firas Hmida [Univ de Rennes I, Temporary Assistant Professor, from Sep 2017] Gabriel Sargent [CNRS, until Aug 2017] Ronan Sicre [CNRS, until Aug 2017] Arnaud Touboulic [CNRS, until Jan 2017]

### **PhD Students**

Rémi Bois [CNRS, until Sep 2017] Ricardo Carlini Sperandio [Univ de Rennes I] Clément Dalloux [CNRS] Mikail Demirdelen [Univ de Rennes I] Cheikh Brahim El Vaigh [Inria, from Oct 2017] Ahmet Iscen [Inria, until Aug 2017] Gregoire Jadi [Univ de Nantes, until Oct 2017] Cédric Maigrot [Univ de Rennes I] Oriane Simeoni [Inria] Vedran Vukotić [INSA Rennes, until Sep 2017] Hanwei Zhang [East China Normal University, Shanghai, from Sep 2017]

### **Post-Doctoral Fellow**

Mateusz Budnik [CNRS, from Apr 2017]

### Visiting Scientist

Gabriel Barbosa Da Fonseca [PUC Minas, Brazil, until Jan 2017]

Administrative Assistant Aurélie Patier [Univ de Rennes I]

## 2. Overall Objectives

### 2.1. Context

Linked media appears today as a major challenge, with numerous potential applications in all areas of multimedia. The strong increase of ubiquitous access to the Internet and the resulting convergence of media on the network open countless opportunities for linked media and reinforce the key role of such a challenge. New applications centered on the notion of linked media are emerging today, such as second screen applications and recommendation services. However, because of the lack of adequate technology, linking related content is mostly deferred to human operators in current applications or to user behavior analysis, e.g., via collaborative filtering, thus indirectly considering the content. This fact severely limits the opportunities offered by a web of media, in terms of creativity, scalability, representativeness and completeness, thus negatively impacting the spread of linked media and the development of innovative services in the Internet of media.

Most of the research effort in automatic multimedia content analysis has been devoted so far to describing and indexing content on which core tasks around information retrieval and recommendation are built to develop multimedia applications. This general philosophy mostly reposes on a vision where documents are considered as isolated entities, i.e., as a basic unit which is indexed or analyzed regardless of other content items and of context. Considering documents in isolation has enabled key progress in content-based analysis and retrieval on a large scale: e.g., design of generic descriptors, efficient techniques for content-based analysis, fast retrieval methodology. But ignoring the links, implicit or explicit, between content items also appears as a rather strong assumption with direct consequences on algorithms and applications, both in terms of performance and in terms of possibilities.

### 2.2. Scientific objectives

LINKMEDIA investigates a number of key issues related to multimedia collections structured with explicit links: Can we discover what characterizes a collection and makes its coherence? Are there repeating motifs that create natural links and which deserve characterization and semantic interpretation? How to explicitly create links from pairwise distances? What structure should a linked collection have? How do we explain the semantic of a link? How explicit links can be used to improve information retrieval? To improve user experience? In this general framework, the global objective of LINKMEDIA is to develop the scientific, methodological and technological foundations facilitating or automating the creation, the description and the exploitation of multimedia collections structured with explicit links. In particular, we target a number of key contributions in the following areas:

• designing efficient methods dedicated to multimedia indexing and unsupervised motif discovery: efficiently comparing content items on a large scale and finding repeating motifs in an unsupervised manner are two key ingredients of multimedia linking based on a low-level representation of the content;

- improving techniques for structuring and semantic description: better description of multimedia content at a semantic—i.e., human interpretable—level, making explicit the implicit structure when it exists, is still required to make the most of multimedia data and to facilitate the creation of links to a precise target at a semantic level;
- designing and experimenting approaches to multimedia content linking and collection structuring: exploiting low-level and semantic content-based proximity to create explicit links within a collection requires specific methodology departing from pairwise comparison and must be confronted with real data;
- studying new paradigms for the exploitation of linked multimedia content as well as new usages: explicit links within media content collections change how such data is processed by machines and ultimately consumed by humans in ways that have yet to be invented and studied.

### 3. Research Program

### 3.1. Scientific background

LINKMEDIA is a multidisciplinary research team, with multimedia data as the main object of study. We are guided by the data and their specificity—semantically interpretable, heterogeneous and multimodal, available in large amounts, unstructured and disconnected—, as well as by the related problems and applications.

With multimedia data at the center, orienting our choices of methods and algorithms and serving as a basis for experimental validation, the team is directly contributing to the following scientific fields:

- multimedia: content-based analysis; multimodal processing and fusion; multimedia applications;
- computer vision: compact description of images; object and event detection;
- natural language processing: topic segmentation; information extraction;
- information retrieval: high-dimensional indexing; approximate k-nn search; efficient set comparison.

LINKMEDIA also takes advantage of advances in the following fields, adapting recent developments to the multimedia area:

- signal processing: image processing; compression;
- machine learning: deep architectures; structured learning; adversarial learning;
- security: data encryption; differential privacy;
- data mining: time series mining and alignment; pattern discovery; knowledge extraction.

### 3.2. Workplan

Research activities in LINKMEDIA are organized along three major lines of research which build upon the scientific domains already mentioned.

### 3.2.1. Unsupervised motif discovery

As an alternative to supervised learning techniques, unsupervised approaches have emerged recently in multimedia with the goal of discovering directly patterns and events of interest from the data, in a totally unsupervised manner. In the absence of prior knowledge on what we are interested in, meaningfulness can be judged based on one of three main criteria: unexpectedness, saliency and recurrence. This last case posits that repeating patterns, known as motifs, are potentially meaningful, leading to recent work on the unsupervised discovery of motifs in multimedia data [56], [54], [55].

LINKMEDIA seeks to *develop unsupervised motif discovery approaches which are both accurate and scalable*. In particular, we consider the discovery of repeating objects in image collections and the discovery of repeated sequences in video and audio streams. Research activities are organized along the following lines:

- developing the scientific basis for scalable motif discovery: sparse histogram representations; efficient co-occurrence counting; geometry and time aware indexing schemes;
- designing and evaluating accurate and scalable motif discovery algorithms applied to a variety of multimedia content: exploiting efficient geometry or time aware matching functions; fast approximate dynamic time warping; symbolic representations of multimedia data, in conjunction with existing symbolic data mining approaches;
- developing methodology for the interpretation, exploitation and evaluation of motif discovery algorithms in various use-cases: image classification; video stream monitoring; transcript-free natural language processing (NLP) for spoken document.

### 3.2.2. Description and structuring

Content-based analysis has received a lot of attention from the early days of multimedia, with an extensive use of supervised machine learning for all modalities [57], [51]. Progress in large scale entity and event recognition in multimedia content has made available general purpose approaches able to learn from very large data sets and performing fairly decently in a large number of cases. Current solutions are however limited to simple, homogeneous, information and can hardly handle structured information such as hierarchical descriptions, tree-structured or nested concepts.

LINKMEDIA aims at *expanding techniques for multimedia content modeling, event detection and structure analysis.* The main transverse research lines that LINKMEDIA will develop are as follows:

- context-aware content description targeting (homogeneous) collections of multimedia data: latent variable discovery; deep feature learning; motif discovery;
- secure description to enable privacy and security aware multimedia content processing: leveraging encryption and obfuscation; exploring adversarial machine learning in a multimedia context; privacy-oriented image processing;
- multilevel modeling with a focus on probabilistic modeling of structured multimodal data: multiple kernels; structured machine learning; conditional random fields.

### 3.2.3. Linking and collection data model

Creating explicit links between media content items has been considered on different occasions, with the goal of seeking and discovering information by browsing, as opposed to information retrieval via ranked lists of relevant documents. Content-based link creation has been initially addressed in the hypertext community for well-structured texts [50] and was recently extended to multimedia content [58], [53], [52]. The problem of organizing collections with links remains mainly unsolved for large heterogeneous collections of unstructured documents, with many issues deserving attention: linking at a fine semantic grain; selecting relevant links; characterizing links; evaluating links; etc.

LINKMEDIA targets pioneering research on media linking by *developing scientific ground, methodology and technology for content-based media linking* directed to applications exploiting rich linked content such as navigation or recommendation. Contributions are concentrated along the following lines:

- algorithmic of linked media for content-based link authoring in multimedia collections: time-aware graph construction; multimodal hypergraphs; large scale k-nn graphs;
- link interpretation and characterization to provide links semantics for interpretability: text alignment; entity linking; intention vs. extension;
- linked media usage and evaluation: information retrieval; summarization; data models for navigation; link prediction.

## 4. Application Domains

### 4.1. Asset management in the entertainement business

Regardless of the ingestion and storage issues, media asset management—archiving, describing and retrieving multimedia content—has turned into a key factor and a huge business for content and service providers. Most content providers, with television channels at the forefront, rely on multimedia asset management systems to annotate, describe, archive and search for content. So do archivists such as the Institut National de l'Audiovisuel, the Nederlands Instituut voor Beeld en Geluid or the British Broadcast Corporation, as well as media monitoring companies, such as Yacast in France. Protecting copyrighted content is another aspect of media asset management.

### 4.2. Multimedia Internet

One of the most visible application domains of linked multimedia content is that of multimedia portals on the Internet. Search engines now offer many features for image and video search. Video sharing sites also feature search engines as well as recommendation capabilities. All news sites provide multimedia content with links between related items. News sites also implement content aggregation, enriching proprietary content with user-generated content and reactions from social networks. Most public search engines and Internet service providers offer news aggregation portals.

### 4.3. Multiscreen TV

The convergence between television and the Internet has accelerated significantly over the past few years, with the democratization of TV on-demand and replay services and the emergence of social TV services and multiscreen applications. These evolutions and the consequently ever growing number of innovative applications offer a unique playground for multimedia technologies. Recommendation plays a major role in connected TV. Enriching multimedia content, with explicit links targeting either multimedia material or knowledge databases, appears as a key feature in this context, at the core of rich TV and second screen applications.

### 4.4. E-learning

On-line courses are rapidly gaining interest with the recent movement for massive open on-line courses (MOOCs). Such courses usually aggregate multimedia material, such as a video of the course with handouts and potentially text books, exercises and other related resources. This setting is very similar to that of the media aggregation sites though in a different domain. Automatically analyzing and describing video and textual content, synchronizing all material available across modalities, creating and characterizing links between related material or between different courses are all necessary features for on-line courses authoring.

## 5. Highlights of the Year

### 5.1. Highlights of the Year

LINKMEDIA organizer of the IEEE Workshop on Information Forensics and Security in Dec. 2017.

### 5.1.1. Awards

Best demo award at ACM Multimedia 2017 for collaborative work within the FUI project NexGenTV.

Best poster award at Advances in Intelligent Data Analysis.

Best paper award with colleagues of IRISA's EXPRESSION team at 24e conférence sur le Traitement Automatique des Langues Naturelles.

BEST PAPERS AWARDS :

[12] MM 2017 - 25th ACM International Conference on Multimedia. O. AHMED, G. SARGENT, F. GARNIER, B. HUET, V. CLAVEAU, L. COUTURIER, R. TRONCY, G. GRAVIER, P. BOUZY, F. LEMÉNOREL.
[28] IDA 2017 - 16th International Symposium on Intelligent Data Analysis. A. LODS, S. MALINOWSKI, R. TAVENARD, L. AMSALEG.

[40] Traitement automatique du langage naturel (TALN). R. QADER, G. LECORVÉ, D. LOLIVE, P. SÉBILLOT.

## 6. New Software and Platforms

### 6.1. Platforms

### 6.1.1. AllGO multimedia web services

Participants: Vincent Claveau, Clément Dalloux, Guillaume Gravier [correspondent], Gabriel Sargent.

Available at http://allgo.irisa.fr, the AllGO platform allows for the easy deployment of the technology developed in the team as web services. Based on the AllGO infrastructure, LINKMEDIA has continued making available a number of web services related to multimedia content analysis. In 2017, we continued our effort towards the interoperability of the services available (silence detection, face detection, text-based fragmentation) and added speaker diarization and negative sentence detection services.

### 7. New Results

### 7.1. Multimedia indexing, Motif and knowledge discovery

### 7.1.1. Towards engineering a web-scale multimedia service: a case study using SPARK Participant: Laurent Amsaleg.

Joint work with Gylfi Þór Guðmundsson (Univ. Reykyavik), Björn Þór Jónsson (Univ. Copenhagen) and Michael J. Franklin (UC Berkeley).

Computing power has now become abundant with multi-core machines, grids and clouds, but it remains a challenge to harness the available power and move towards gracefully handling web-scale datasets. Several researchers have used automatically distributed computing frameworks, notably Hadoop and Spark, for processing multimedia material, but mostly using small collections on small clusters. We describe the engineering process for a prototype of a (near) web-scale multimedia service using the Spark framework running on the AWS cloud service. We present experimental results using up to 43 billion SIFT feature vectors from the public YFCC 100M collection, making this the largest high-dimensional feature vector collection reported in the literature. The design of the prototype and performance results demonstrate both the flexibility and scalability of the Spark framework for implementing multimedia services.

## 7.1.2. On competitiveness of nearest-neighbor based music classification: a methodological critique

#### Participant: Laurent Amsaleg.

Joint work with Haukur Pálmasson, Björn Þór Jónsson (Univ. Copenhagen), Markus Schedl (Johannes Kepler University), Peter Knees (TU Wien). The traditional role of nearest-neighbor classification in music classification research is that of a straw man opponent for the learning approach of the hour. Recent work in high-dimensional indexing has shown that approximate nearest-neighbor algorithms are extremely scalable, yielding results of reasonable quality from billions of high-dimensional features. With such efficient large-scale classifiers, the traditional music classification methodology of reducing both feature dimensionality and feature quantity is incorrect; instead the approximate nearest-neighbor classifier should be given an extensive data collection to work with. We present a case study, using a well-known MIR classification benchmark with well-known music features, which shows that a simple nearest-neighbor classifier performs very competitively when given ample data. In this position paper, we therefore argue that nearest-neighbor classification has been treated unfairly in the literature and may be much more competitive than previously thought [30].

### 7.1.3. Unsupervised part learning for visual recognition

Participants: Ronan Sicre, Yannis Avrithis, Ewa Kijak.

### Joint work with Frederic Jurie (Univ. Caen).

Part-based image classification aims at representing categories by small sets of learned discriminative parts, upon which an image representation is built. Considered as a promising avenue a decade ago, this direction has been neglected since the advent of deep neural networks. In this context, the work proposed here brings two contributions: first, this work proceeds one step further compared to recent part-based models (PBM), focusing on how to learn parts without using any labeled data. Instead of learning a set of parts per class, as generally performed in the PBM literature, the proposed approach constructs a partition of a given set of images into visually similar groups, and subsequently learns a set of discriminative parts per group in a fully unsupervised fashion. This strategy opens the door to the use of PBM in new applications where labeled data are typically not available, such as instance-based image retrieval. Second , we show that despite the recent success of end-to-end models, explicit part learning can still boost classification performance. We experimentally show that our learned parts can help building efficient image representations, which outperform state-of-the art deep convolutional neural networks on both classification and retrieval tasks [32].

### 7.1.4. Automatic discovery of discriminative parts as a quadratic assignment problem Participants: Ronan Sicre, Yannis Avrithis, Teddy Furon, Ewa Kijak.

#### Joint work with Julien Rabin and Frédéric Jurie (Univ. Caen).

Part-based image classification consists in representing categories by small sets of discriminative parts upon which a representation of the images is built. This piece of work addresses the question of how to automatically learn such parts from a set of labeled training images. We propose to cast the training of parts as a quadratic assignment problem in which optimal correspondences between image regions and parts are automatically learned. We analyze different assignment strategies and thoroughly evaluates them on two public datasets: Willow actions and MIT 67 scenes [45].

### 7.1.5. Learning DTW-preserving shapelets

Participants: Laurent Amsaleg, Arnaud Lods, Simon Malinowski.

#### Joint work with Romain Tavenard (Univ. Rennes 2).

Dynamic time warping (DTW) is one of the best similarity measures for time series, and it has extensively been used in retrieval, classification or mining applications. It is a costly measure, and applying it to numerous and/or very long times series is difficult in practice. Recently, shapelet transform (ST) proved to enable accurate supervised classification of time series. ST learns small subsequences that well discriminate classes, and transforms the time series into vectors lying in a metric space. We adopt the ST framework in a novel way: we focus on learning, without class label information, shapelets such that Euclidean distances in the ST-space approximate well the true DTW. Our approach leads to an ubiquitous representation of time series in a metric space, where any machine learning method (supervised or unsupervised) and indexing system can operate efficiently [28].

## 7.1.6. Tag propagation approaches within speaking face graphs for multimodal person discovery

Participants: Guillaume Gravier, Gabriel Sargent, Ronan Sicre.

Joint work with Gabriel Barbosa Da Fonseca, Izabela Lyon Freire, Zenilton Patrocinio Jr and Silvio Jamil F. Guimaraes (PUC Minas, Brazil)

The indexing of broadcast TV archives is a current problem in multimedia research. As the size of these databases grows continuously, meaningful features are needed to describe and connect their elements efficiently, such as the identification of speaking faces. In this context, we focused on two approaches for unsupervised person discovery. Initial tagging of speaking faces is provided by an OCR-based method, and these tags propagate through a graph model based on audiovisual relations between speaking faces. Two propagation methods are proposed, one based on random walks and the other based on a hierarchical approach. To better evaluate their performances, these methods were compared with two graph clustering baselines. We also study the impact of different modality fusions on the graph-based tag propagation scenario. From a quantitative analysis, we observed that the graph propagation techniques always outperform the baselines. Among all compared strategies, the methods based on hierarchical propagation with late fusion and random walk with score-fusion obtained the highest MAP values. Finally, even though these two methods produce highly equivalent results according to Kappa coefficient, the random walk method performs better according to a paired t-test, and the computing time for the hierarchical propagation is more than 4 times lower than the one for the random walk propagation [22].

The tag propagation results were included in a large-scale comparison of systems for person discovery in broadcast videos resulting from the MediaEval 2016 international benchmark [27].

### 7.2. Multimedia content description and structuring

## 7.2.1. The vulnerability of learning to adversarial perturbation increases with intrinsic dimensionality

Participant: Laurent Amsaleg.

Joint work with James Bailey, Dominique Barbe, Sarah Erfani, Michael Houle, Vinh Nguyen amd Miloš Radovanovic.

Recent research has shown that machine learning systems, including state-of-the-art deep neural networks, are vulnerable to adversarial attacks. By adding to the input object an imperceptible amount of adversarial noise, it is highly likely that the classifier can be tricked into assigning the modified object to any desired class. It has also been observed that these adversarial samples generalize well across models. A complete understanding of the nature of adversarial samples has not yet emerged. Towards this goal, we present a novel theoretical result formally linking the adversarial vulnerability of learning to the intrinsic dimensionality of the data. In particular, our investigation establishes that as the local intrinsic dimensionality (LID) increases, 1-NN classifiers become increasingly prone to being subverted. We show that in expectation, a k-nearest neighbor of a test point can be transformed into its 1-nearest neighbor by adding an amount of noise that diminishes as the LID increases. We also provide an experimental validation of the impact of LID on adversarial perturbation for both synthetic and real data, and discuss the implications of our result for general classifiers [13].

### 7.2.2. Efficient temporal kernels between feature sets for time series classification Participant: Simon Malinowski.

Joint work with Romain Tavenard, Adeline Bailly, Louis Chapel (Univ. Rennes 2), Benjamin Bustos and Heider Sanchez (Univ. of Chile).

In the time-series classification context, the majority of the most accurate core methods are based on the bag-of-words framework, in which sets of local features are first extracted from time series. A dictionary of words is then learned and each time series is finally represented by a histogram of word occurrences. This representation induces a loss of information due to the quantization of features into words as all the time series are represented using the same fixed dictionary. In order to overcome this issue, we have designed a kernel operating directly on sets of features. Then, we have extended it to a time-compliant kernel that allows one to take into account the temporal information. We applied this kernel in the time series classification context. Proposed kernel has a quadratic complexity with the size of input feature sets, which is problematic when dealing with long time series. However, we have shown that kernel approximation techniques can be used to define a good trade-off between accuracy and complexity. We experimentally demonstrated that the proposed kernel can significantly improve the performance of time series classification algorithms based on bag-of-words [33].

### 7.2.3. Tampering detection and localization in images from social networks

Participants: Cédric Maigrot, Ewa Kijak, Vincent Claveau.

Verifying the authenticity of an image broadcast on social networks is crucial to limit the dissemination of false information. In this work, we aim to provide information about tampering localisation on such images, in order to help either the user or automatic methods to discriminate truth from falsehood. These images may have been subjected to a large number of possible forgeries, which calls for the use of generic methods. Image forensics methods based on local features have proven to be effective for the specific case of copy-move forgery. By taking advantage of the number of images available on the internet, we propose a generic system based on image retrieval, and image comparison based on local features to localise any kind of tampering in images from social networks.

Images from social media are likely to have undergone a large variety of modifications, some being malicious, and some not. The proposed approach is evaluated on three dedicated datasets containing a variety of representative tamperings in images from social media, with difficult examples. This allows an analysis of the local-features approaches behavior in this context. The method is further compared to several state-of-the-art methods and proves to be superior. Finally, we propose a classification step to discriminate malicious modifications from the non-malicious ones.

We have also built and made publicly available a large and challenging adapted database of real case images for evaluation [29].

### 7.2.4. Identity documents classification as an image classification problem

### Participants: Ronan Sicre, Teddy Furon.

Joint work with Ahmad Montaser Awal and Nabil Ghanni (AriadNext).

This works studies the classification of images of identification documents. More specifically, we address the classification of documents composed of few textual information and complex background (such as identity documents). Unlike most existing systems, the proposed approach simultaneously locates the document and recognizes its class. The latter is defined by the document nature (passport, ID, etc.), emission country, version, and the visible side (main or back). This task is very challenging due to unconstrained capturing conditions, sparse textual information, and varying components that are irrelevant to the classification, e.g. photo, names, address, etc. First, a base of document models is created from reference images.

This problem is critical in various security context where proposed system must offer high performances. We address this challenge as an image classification problem, which has received a large attention from the scientific community. We show that training images are not necessary and only one reference image is enough to create a document model. Then, the query image is matched against all models in the base. Unknown documents are rejected using an estimated quality based on the extracted document. The matching process is optimized to guarantee an execution time independent from the number of document models. Once the document model is found, a more accurate matching is performed to locate the document and facilitate information extraction. Our system is evaluated on several datasets with up to 3042 real documents (representing 64 classes) achieving an accuracy of 96.6 % in [14].

In a second step, several methods are evaluated and we report results allowing a better understanding of the specificity of identification documents. We are especially interested in deep learning approaches, showing good transfer capabilities and high performances [44], [49].

### 7.2.5. Sentiment analysis

Participants: Vincent Claveau, Christian Raymond.

In the framework of the NexGenTV project, we have participated to the text-mining challenge DeFT about sentiment analysis. We have proposed methods for the identification of figurative language (irony, humor...), and for the classification of figurative and non-figurative tweets according to their polarity. For these tasks, we explore the use of three methods of increasing complexity: i) k-nearest neighbors with information retrieval based techniques, ii) boosting of decision trees, iii) recurrent neural networks [36]. It allows us to evaluate the precise interest of each of our approach and the data representation that they use: bag-of-words for the first one, n-grams for the second and word embedding for the latest.

### 7.3. Content-based information retrieval

## 7.3.1. Efficient diffusion on region manifolds: recovering small objects with compact CNN representations

Participants: Yannis Avrithis, Teddy Furon, Ahmet Iscen.

Joint work with Giorgos Tolias and Ondrej Chum (Technical University of Prague).

Query expansion is a popular method to improve the quality of image retrieval with both conventional and CNN representations. It has been so far limited to global image similarity. This work focuses on diffusion, a mechanism that captures the image manifold in the feature space. The diffusion is carried out on descriptors of overlapping image regions rather than on a global image descriptor like in previous approaches. An efficient off-line stage allows optional reduction in the number of stored regions. In the on-line stage, the proposed handling of unseen queries in the indexing stage removes additional computation to adjust the precomputed data. We perform diffusion through a sparse linear system solver, yielding practical query times well below one second. Experimentally, we observe a significant boost in performance of image retrieval with compact CNN descriptors on standard benchmarks, especially when the query object covers only a small part of the image. Small objects have been a common failure case of CNN-based retrieval [25].

### 7.3.2. Panorama to panorama matching for location recognition Participants: Yannis Avrithis, Teddy Furon, Ahmet Iscen.

### Joint work with Giorgos Tolias and Ondrej Chum (Technical University of Prague).

Location recognition is commonly treated as visual instance retrieval on "street view" imagery. The dataset items and queries are panoramic views, i.e., groups of images taken at a single location. This work introduces a novel panorama-to-panorama matching process, either by aggregating features of individual images in a group or by explicitly constructing a larger panorama. In either case, multiple views are used as queries. We reach near perfect location recognition on a standard benchmark with only four query views [26].

### 7.3.3. Memory vectors for similarity search in high-dimensional spaces Participants: Teddy Furon, Ahmet Iscen.

Joint work with Vincent Gripon (IMT Atlantique), Michael Rabbat (Mc Gill University), and Hervé Jégou (Facebook AI Research).

We study an indexing architecture to store and search in a database of high-dimensional vectors from the perspective of statistical signal processing and decision theory. This architecture is composed of several memory units, each of which summarizes a fraction of the database by a single representative vector. The potential similarity of the query to one of the vectors stored in the memory unit is gauged by a simple correlation with the memory unit's representative vector. This representative optimizes the test of the following hypothesis: the query is independent from any vector in the memory unit vs. the query is a simple perturbation of one of the stored vectors. Compared to exhaustive search, our approach finds the most similar database vectors significantly faster without a noticeable reduction in search quality. Interestingly, the reduction of complexity is provably better in high-dimensional spaces. We empirically demonstrate its practical interest in a large-scale image search scenario with off-the-shelf state-of-the-art descriptors [6].

### 7.3.4. Exploiting multimodality in video hyperlinking to improve target diversity

**Participants:** Rémi Bois, Guillaume Gravier, Christian Raymond, Pascale Sébillot, Ronan Sicre, Vedran Vukotić.

Video hyperlinking is the process of creating links within a collection of videos to help navigation and information seeking. Starting from a given set of video segments, called anchors, a set of related segments, called targets, must be provided. In past years, a number of content-based approaches have been proposed with good results obtained by searching for target segments that are very similar to the anchor in terms of content and information. Unfortunately, relevance has been obtained to the expense of diversity. In this paper, we study multimodal approaches and their ability to provide a set of diverse yet relevant targets. We compare two recently introduced cross-modal approaches, namely, deep auto-encoders and bimodal LDA, and experimentally show that both provide significantly more diverse targets than a state-of-the-art baseline. Bimodal autoencoders offer the best trade-off between relevance and diversity, with bimodal LDA exhibiting slightly more diverse targets at a lower precision [17].

## 7.3.5. Generative adversarial networks for multimodal representation learning in video hyperlinking

Participants: Guillaume Gravier, Christian Raymond, Vedran Vukotić.

Continuous multimodal representations suitable for multimodal information retrieval are usually obtained with methods that heavily rely on multimodal autoencoders. In video hyperlinking, a task that aims at retrieving video segments, the state of the art is a variation of two interlocked networks working in opposing directions. These systems provide good multimodal embeddings and are also capable of translating from one representation space to the other. Operating on representation spaces, they lack the ability to operate in the original spaces (text or image), which makes it difficult to visualize the crossmodal function, and do not generalize well to unseen data. Recently, generative adversarial networks (GANs) have gained popularity and have been used for generating realistic synthetic data and for obtaining high-level, single-modal latent representations. We show that GANs can be used for multimodal representation learning and that they provide multimodal representations that are superior to representations obtained with multimodal autoencoders. Additionally, we illustrate the ability of visualizing crossmodal translations that can provide human-interpretable insights on learned GAN-based video hyperlinking models [35].

### 7.4. Linking, navigation and analytics

### 7.4.1. Providing real-time insight during political debates in a second screen application Participants: Vincent Claveau, Guillaume Gravier, Gabriel Sargent.

Joint work with Institut Eurecom, Wildmoka and AVISTO Telecom in the framework of the FUI project NexGenTV.

Second screen applications are becoming key for broadcasters exploiting the convergence of TV and Internet. Authoring such applications however remains costly. Within the NexGenTV project, we developed a second screen authoring application that leverages multimedia content analytics and social media monitoring. A back-office is dedicated to easy and fast content ingestion, segmentation, description and enrichment with links to entities and related content. From the back-end, broadcasters can push enriched content to frontend applications providing customers with highlights, entity and content links, overviews of social network, etc. The demonstration operates on political debates ingested during the 2017 French presidential election,

http://www.nexgentv.fr/communication/events/discover-our-new-politics-debates-live-video-edition

### 7.4.2. Information extraction in clinical documents

enabling insights on the debates [12].

Participants: Clément Dalloux, Vincent Claveau.

Joint work with Claudia Moro (Pontifícia Universidade Católica do Paraná, Brazil) and Natalia Grabar (Univ. Lille)

Extracting fine-grained information from clinical texts is a keystone for numerous medical applications. For instance, in clinical trial protocols eligibility criteria are expressed through texts in an unstructured way. This year, we have developed an annotated corpus of clinical trials and made it available to the community. Based on this corpus, we proposed automatic methods to extract numerical information [20] and to handle the variation of the units used [43]. In such medical applications, detecting negation, uncertainty, and the scope on which they apply is important. Thus, we have also developed an annotated corpus, made it available to the community, and we have proposed an automatic tool based on recurrent neural networks [37], [41] and made it available as a web service.

### 7.4.3. Semi-supervision for information extraction

Participants: Vincent Claveau, Ewa Kijak.

Many NLP problems are tackled as supervised machine learning tasks. Consequently, the cost of the expertise needed to annotate the examples is a widespread issue. Active learning offers a framework to that issue, allowing to control the annotation cost while maximizing the classifier performance, but it relies on the key step of choosing which example will be proposed to the expert. This year, we examined and proposed such selection strategies in the specific case of conditional random fields (CRF) which are largely used in NLP. On the one hand, we proposed a simple method to correct a bias of some state-of-the-art selection techniques. On the other hand, we built an original approach to select the examples, based on the respect of proportions in the datasets. These contributions were validated over a large range of experiments implying several datasets and tasks, including named entity recognition, chunking, phonetization, word sense disambiguation [19].

### 7.4.4. Linking multimedia content for efficient news browsing via explorable news graphs

Participants: Rémi Bois, Guillaume Gravier, Pascale Sébillot.

Joint work with Maxime Robert, Éric Jamet (Univ. Rennes 2) and Emmanuel Morin (Univ. Nantes) in the framework of the CominLabs project Linking Media in Acceptable Hypergraphs.

As the amount of news information available online grows, media professionals are in need of advanced tools to explore the information surrounding specific events before writing their own piece of news, e.g., adding context and insight. While many tools exist to extract information from large datasets, they do not offer an easy way to gain insight from a news collection by browsing, going from article to article and viewing unaltered original content. Such browsing tools require the creation of rich underlying structures such as graph representations. These representations can be further enhanced by typing links that connect nodes, in order to inform the user on the nature of their relation. We propose an efficient way to generate links between news items in order to obtain an easily navigable graph, and enrich this graph by automatically typing created links. User evaluations are conducted on real-world data in order to assess for the interest of both the graph representation and link typing in a press reviewing task, showing a significant improvement compared to classical search engines [15], [16].

### 7.4.5. Multimodal detection of fake news

Participants: Vincent Claveau, Cédric Maigrot, Ewa Kijak.

Social networks make it possible to share rapidly and massively information, including fake news, hoaxes or rumors. Following our previous work in the frame of the Verification Multimedia Use task of Mediaeval 2016, we have explored the use of multimodal clues to detect fake news in social networks [38]. This year, we have studied the interest of combining and merging many approaches developed by the MediaEval participants in order to evaluate the predictive power of each modality. We have proposed several fusion strategies making the most of their potential complementarity [39].

### 7.5. Miscellaneous

In parallel with mainstream research activities, LINKMEDIA has a number of contributions in other domains based on the expertise of the team members.

## 7.5.1. One-step time-dependent future video frame prediction with a convolutional encoder-decoder neural network

Participants: Guillaume Gravier, Christian Raymond, Vedran Vukotić.

Joint work with Silvia-Laura Pintea and Jan Van Gemert (TU Delft, The Netherlands).

There is an inherent need for autonomous cars, drones, and other robots to have a notion of how their environment behaves and to anticipate changes in the near future. In this work, we focus on anticipating future appearance given the current frame of a video. Existing work focuses on either predicting the future appearance as the next frame of a video, or predicting future motion as optical flow or motion trajectories starting from a single video frame. This work stretches the ability of convolutional neural networks (CNNs) to predict an anticipation of appearance at an arbitrarily given future time, not necessarily the next video frame. We condition our predicted future appearance on a continuous time variable that allows us to anticipate future frames at a given temporal distance, directly from the input video frame. We show that CNNs can learn an intrinsic representation of typical appearance changes over time and successfully generate realistic predictions at a deliberate time difference in the near future [34].

### 7.5.2. About zero bit watermarking error exponents

### Participant: Teddy Furon.

This work aims to motivate more research works on the design of zero-bit watermarking schemes by showing an upper bound of the performances that known solutions failed to reach. To this end, an upper bound of error exponent characteristic is derived by translating Costa's rationale to zero-bit watermarking with side information. Three schemes are then considered: the dual-cone detection region originally proposed by Cox et al. and improved in Merhav et al. papers, ISS (Improved Spread Spectrum), and ZATT (Zero Attraction). It turns out that in certain conditions the latter performs better than the first one, which questions the optimality claimed Merhav et al. Nevertheless, the main conclusion is that these schemes are in general far away from the upper bound in the region of practical interest [23].

### 8. Bilateral Contracts and Grants with Industry

### 8.1. Bilateral Contracts with Industry

Quai des Apps: one-shot 2-day contract for scientific counseling on visual image retrieval.

### 9. Partnerships and Cooperations

### 9.1. Regional Initiatives

### 9.1.1. CominLabs Project Linking Media in Acceptable Hypergraphs (LIMAH)

Participants: Rémi Bois, Vincent Claveau, Guillaume Gravier, Pascale Sébillot, Arnaud Touboulic.

Duration: 4 years, started in April 2014

Partners: Telecom Bretagne (IODE), Univ. Rennes II (CRPCC, PREFics), Univ. Nantes (LINA/TAL) URL: http://limah.irisa.fr

LIMAH aims at exploring hypergraph structures for multimedia collections, instantiating actual links reflecting particular content-based proximity—similar content, thematic proximity, opinion expressed, answer to a question, etc. Exploiting and developing further techniques targeting pairwise comparison of multimedia contents from an NLP perspective, LIMAH addresses two key issues: How to automatically build from a collection of documents an hypergraph, i.e., a graph combining edges of different natures, which provides exploitable links in selected use cases? How collections with explicit links modify usage of multimedia data in all aspects, from a technology point of view as well as from a user point of view? LIMAH studies hypergraph authoring and acceptability taking a multidisciplinary approach mixing ICT, law, information and communication science as well as cognitive and ergonomy psychology.

### 9.1.2. CominLabs Project BigCLIN

Participants: Vincent Claveau, Ewa Kijak, Clément Dalloux.

Duration: 3 years, started in September 2016 Partners: STL-CNRS, Inserm/CHU Rennes, Inria URL: http://www.bigclin.cominlabs.ueb.eu

Data collected or produced during clinical care process can be exploited at different levels and across different domains. Yet, a well-known challenge for secondary use of health big data is that much of detailed patient information is embedded in narrative text, mostly stored as unstructured data. The project proposes to address the essential needs when reusing unstructured clinical data at a large scale. We propose to develop new clinical records representation relying on fine-grained semantic annotation thanks to new NLP tools dedicated to French clinical narratives. To efficiently map this added semantic information to existing structured data for further analysis at big scale, the project also addresses distributed systems issues: scalability, management of uncertain data and privacy, stream processing at runtime, etc.

### 9.2. National Initiatives

### 9.2.1. ANR Project IDFRAud

Participant: Teddy Furon.

### Duration: 3 years, started in Feb. 2015 Partners: AriadNext, IRCGN, École Nationale Supérieure de Police

The IDFRAud project consists in proposing an automatic solution for ID analysis and integrity verification. Our ID analysis goes through three processes: classification, text extraction and ID verification. The three processes rely on a set of rules that are externalized in formal manner in order to allow easy management and evolving capabilities. This leads us to the ID knowledge management module. Finally, IDFRAud addresses the forensic link detection problem and to propose an automatic analysis engine that can be continuously applied on the detected fraud ID database. Cluster analysis methods are used to discover relations between false IDs in their multidimensional feature space. This pattern extraction module will be coupled with a suitable visualization mechanism in order to facilitate the comprehension and the analysis of extracted groups of interlinked fraud cases.

### 9.2.2. FUI 19 NexGenTV

Participants: Vincent Claveau, Guillaume Gravier, Ewa Kijak, Gabriel Sargent, Ronan Sicre.

### Duration: 2.5 years, started in May 2015

Partners: Eurecom, Avisto Telecom, Wildmoka, Envivio-Ericsson

Television is undergoing a revolution, moving from the TV screen to multiple screens. Today's user watches TV and, at the same time, browses the web on a tablet, sends SMS, posts comments on social networks, searches for complementary information on the program, etc. Facing this situation, NexGen-TV aims at developing a generic solution for the enrichment, the linking and the retrieval of video content targeting the cost-cutting edition of second screen and multiscreen applications for broadcast TV. The main outcome of the project will be a software platform to aggregate and distribute video content via a second-screen edition interface connected to social media. The curation interface will primarily make use of multimedia and social media content segmentation, description, linking and retrieval. Multiscreen applications will be developed on various domaine, e.g., sports, news.

### 9.2.3. Inria Project Lab Knowledge-driven data and content collaborative analytics (iCODA)

**Participants:** Laurent Amsaleg, Vincent Claveau, Cheikh Brahim El Vaigh, Guillaume Gravier, Pascale Sébillot.

Duration: 4.5 years, started in April 2017

Partners: Inria project-teams Linkmedia, CEDAR, GraphIK and ILDA, with Ouest France, Le Monde and AFP

One of today's major issues in data science is the design of algorithms that allow analysts to efficiently infer useful information and knowledge by collaboratively inspecting heterogeneous information sources, from structured data to unstructured content. Taking data journalism as an emblematic use-case, the goal of the project is to develop the scientific and technological foundations for knowledge- mediated user-in-the-loop collaborative data analytics on heterogeneous information sources, and to demonstrate the effectiveness of the approach in realistic, high-visibility use-cases. The project stands at the crossroad of multiple research fields—content analysis, data management, knowledge representation, visualization—that span multiple Inria themes, and counts on a club of major press partners to define usage scenarios, provide data and demonstrate achievements.

### 9.3. European Initiatives

### 9.3.1. CHIST ERA ID\_IOT

Participant: Teddy Furon.

Duration: 3 years, started in Oct. 2016 Partners: Eindhoven Univ. of Technology, Univ. of Geneva

The IoT will contain a huge number of devices and objects that have very low or nonexistent processing and communication resources, coupled to a small number of high-power devices. The weakest devices, which are most ubiquitous, will not be able to authenticate themselves using cryptographic methods. This project addresses these issues using physical unclonable functions (PUFs). PUFs, and especially quantum readout PUFs, are ideally suited to the IoT setting because they allow for the authentication and identification of physical objects without requiring any crypto or storage of secret information.

Furthermore, we foresee that back-end systems will not be able to provide security and privacy via cryptographic primitives due to the sheer number of IoT devices. Our plan is to address these problems using privacypreserving database structures and algorithms with good scaling behaviour. Approximate nearest neighbour (ANN) search algorithms, which have remarkably good scaling behaviour, have recently become highly efficient, but do not yet have the right security properties and have not yet been applied to PUF data. Summarised in a nutshell, the project aims to improve the theory and practice of technologies such as PUFs and ANN search in the context of generic IoT authentication and identification scenarios.

### 9.3.2. Collaborations with Major European Organizations

Big Data Value Association (BDVA): LINKMEDIA is a co-founder and co-leader of the media group (TF7) within BDVA

### 9.4. International Initiatives

### 9.4.1. Inria International Partners

9.4.1.1. Informal International Partners

- National Institute for Informatics, Japan
- University of Amsterdam, The Netherlands
- Czech Technical University, Czech Republic
- Katholieke Universiteit Leuven, Belgium

### 9.4.2. Participation in Other International Programs

- CNRS CONFAP FIGTEM
  - Title: Fine-grained text-mining for clinical trials
  - International Partner (Institution Laboratory Researcher): Pontifícia Universidade Católica do Paraná - Health Informatics dept, Claudia Moro

FIGTEM aims at developing natural language processing methods, including information extraction and indexing, dedicated to the clinical trial domain. The goal is to populate a formal representation of patients (via their electronic patient records) and clinical trial data in different languages (French, English, Portuguese).

- Jan. 2016 – Dec. 2018

### 9.5. International Research Visitors

### 9.5.1. Visits of International Scientists

### **Giorgos Tolias**

Date: Sept. 2017 (1 week)

Institution: Czech Technical University, Czech Republic

### Vincent Oria

Date: July 2017 (2 weeks)

Institution: New Jersey Institute of Technology, Newark, USA

#### Michael Houle

Date: July 2017 (2 weeks)

Institution: National Institute of Informatics, Tokyo, Japan

### 9.5.1.1. Internships

Gabriel B. de Fonseca

Date: Nov. 2016 - Jan. 2017

Institution: PUC Minas, Brazil

### 9.5.2. Visits to International Teams

Laurent Amsaleg

Date: Oct. 2017 (2 days) Institution: New Jersey Institute of Technology, Newark, USA Laurent Amsaleg Date: May. 2017 (1 week) Institution: East China Normal University, Shanghai, PRC Clément Dalloux Date: Nov.-Dec 2017 (1 month) Institution: Pontifícia Universidade Católica do Paraná, Brazil Guillaume Gravier Date: Nov. 2017 (1 week) Institution: Universidad de Chile, Santiago, Chile Guillaume Gravier Date: May. 2017 (1 week) Institution: East China Normal University, Shanghai, PRC

### **10.** Dissemination

### **10.1. Promoting Scientific Activities**

### 10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

Laurent Amsaleg was appointed general co-chair of ACM Intl. Conf. on Multimedia 2019.

Laurent Amsaleg was general chair of Intl. Conf. on Multimedia Modeling 2017.

Laurent Amsaleg and Vincent Claveau were general co-chairs of the Workshop on Computational Journalism collocated with EGC 2017.

Teddy Furon was general co-chair of IEEE Intl. Workshop on Information Forensics and Security 2017.

Teddy Furon was a co-chair of the GdR-ISIS national workshop "Comment concilier Big Data, identification des personnes, traçabilité des contenus et respect de la vie privée ?", 2017.

Ewa Kijak and Vincent Claveau organized and chaired a special session on Social Networks and User-Generated Content Verification at the IEEE Intl. Workshop on Information Forensics and Security.

10.1.1.2. Member of the Organizing Committees

Guillaume Gravier was special session chair at Intl. Conf. on Multimedia Modeling 2017.

Guillaume Gravier was nominated special session co-chair of IEEE Intl. Workshop on Content-Based Multimedia Indexing 2018.

### 10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

Yannis Avrithis was area chair of the ACM Intl. Conf. on Multimedia 2017.

Yannis Avrithis was area chair of the European Signal Processing Conf 2017.

Guillaume Gravier was apppointed area chair of ACM Intl. Conf. on Multimedia 2019.

10.1.2.2. Member of Conference Program Committees

Laurent Amsaleg was a PC member of: Base de Données AvancéesA; IEEE Intl. Workshop on Content-Based Multimedia Indexing, IEEE Intl. Conf. on Multimedia and Exhibition; Intl. Conf. on Similarity Search and Applications; ACM Intl. Conf. on Multimedia Retrieval.

Yannis Avrithis was a PC member of: IEEE Conf. on Computer Vision and Pattern Recognition; Intl. Conf. on Computer Vision; ACM Intl. Conf. on Multimedia; European Signal Processing Conf.; IEEE Intl. Conf. on Multimedia and Expo; IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing.

Vincent Claveau was a PC member of: ACL demonstration track; EMNLP demonstration track; International Semantic Web Conference; Intl. Conf. on Multimedia Modeling; IEEE Intl. Workshop on Information Forensics and Security; Conférence en Recherche d'Information et Applications; TextMine workshop.

Teddy Furon was a PC member of ACM Information Hiding and Multimedia Security.

Guillaume Gravier was a PC member of: ACM Intl. Conf. on Multimedia; IEEE Intl. Conf. on Multimedia and Exhibition; ACM Intl. Conf. on Multimedia Retrieval; Annual Conf. of the Intl. Speech Communication Association; European Conf. on Information Retrieval; IEEE Intl. Conf. on Acoustics, Speech and Signal Processing; IEEE Intl. Symposium on Multimedia; Multimedia Modeling Conf.

Ewa Kijak was a PC member of: ACM Intl. Conf. on Multimedia; ACM Intl. Conf. on Multimedia Retrieval; Intl. Workshop on Content-Based Multimedia Indexing; IEEE Intl. Workshop on Information Forensics and Security.

Christian Raymond was a PC member of: Annual Conf. of the Intl. Speech Communication Association; IEEE Intl. Conf. on Acoustics Speech and Signal Processing; IEEE Intl. Conf. on Machine Learning And Applications; Conf. en Traitement Automatique des Langues Naturelles.

Pascale Sébillot was a PC member of: European Conf. on Information Retrieval; Annual Meeting of the Association for Computational Linguistics; Intl. Joint Conf. on Artificial Intelligence; Intl. Conf. on Multimedia Modeling; Conf. Traitement Automatique des Langues Naturelles; Traitement Automatique du Langage et Analyse de Documents.

### 10.1.2.3. Reviewer

Ewa Kijak reviewed for ACM Intl. Conf. on Multimedia.

### 10.1.3. Journal

#### 10.1.3.1. Member of the Editorial Boards

Laurent Amsaleg was guest editor for the special issue of the Multimedia tools and applications Journal, best papers from the MMM2017 conference.

Vincent Claveau is editor of the journal Recherche d'Information, Document, Web Sémantique

Vincent Claveau is member of the editoral board of the journal Traitement Automatique des Langues.

Guillaume Gravier is associate editor of IEEE Trans. on Multimedia.

Guillaume Gravier was editor of the Working Notes Proc. of the MediaEval Multimedia Benchmark.

Christian Raymond is member of the editorial board of the electronic Journal Discours.

Pascale Sébillot is editor of the Journal Traitement Automatique des Langues.

Pascale Sébillot is member of the editorial board of the Journal Traitement Automatique des Langues.

### 10.1.3.2. Reviewer - Reviewing Activities

Laurent Amsaleg was a reviewer for: IEEE Trans. on Circuits and Systems for Video Technology; IEEE Trans. on Multimedia, Pattern Recognition Letters; IEEE Trans. on Information Forensics & Security; IEEE Trans. on Image Processing; IEEE Trans. on Information Systems; IEEE Trans. on Knowledge and Data Engineering.

Yannis Avrithis was a reviewer for Multimedia Tools and Applications, Elsevier Neurocomputing.

Vincent Claveau reviewed for reviewed for Multimedia Tools and Applications, Elsevier Information Sciences, Traitement Automatique des Langues.

Teddy Furon reviewed for IEEE Trans. on Information Forensics and Security, IEEE Trans. on Image Processing, and IEEE Trans. on Signal Processing.

Guillaume Gravier reviewed for Information Science, Multimedia Tools and Applications, IEEE Trans. on Multimedia.

Christian Raymond was a reviewer for Multimedia Tools and Applications.

Pascale Sébillot reviewed for Traitement Automatique des Langues.

### 10.1.4. Invited Talks

Laurent Amsaleg gave an invited talk at East China Normal University, Shanghai.

Yannis Avrithis gave an invited talk at Deep Learning in Vision, Language, and Multimedia Workshop on September 25 in Rennes.

Vincent Claveau gave an invited talk at Intelligence artificielle et Recherche d'Information : Journée commune AFIA - ARIA.

Guillaume Gravier gave an invited talk at Katholieke Universiteit Leuven.

Guillaume Gravier gave an invited talk at Universidad de Chile, Santiago de Chile.

Guillaume Gravier gave an invited talk at East China Normal University, Shanghai.

Pascale Sébillot gave an invited tutorial at symposium on Propriété intellectuelle et données dans l'environnement numérique, Rennes.

### 10.1.5. Leadership within the Scientific Community

Laurent Amsaleg is a member of the Steering Committee of SISAP for the 2016–2020 term.

Vincent Claveau is finance head of the Association pour la Recherche d'Informations et ses Applications (ARIA).

Guillaume Gravier is co-founder and general chair of the ISCA SIG Speech, Language and Audio in Multimedia.

Guillaume Gravier is member of the Community Council of the Mediaeval Multimedia Evaluation series.

Guillaume Gravier was president of the Scientific Evaluation Committee of the National Research Agency for the theme 'Knowledge, data, content, big data, HPC, simulation' up to July 2017.

Since September 2017, Guillaume Gravier is president of the Scientific Evaluation Committee of the National Research Agency for the theme 'Knowledge, data, content, big data - AI'.

Pascale Sébillot is a member of the permanent steering committee of Conf. Francophone en Traitement Automatique des Langues Naturelles.

### 10.1.6. Scientific Expertise

Laurent Amsaleg was an expert evaluator for the French National Research Agency (ANR).

Yannis Avrithis was an expert evaluator for the French National Research Agency (ANR).

Vincent Claveau was an expert for was an expert for the Idex UCA-Jedi

Teddy Furon is scientific adviser for the company LAMARK.

### 10.1.7. Research Administration

Vincent Claveau is deputy head of the GdR MaDICS, a CNRS inter-lab initiative to promote research about Big Data and Data Science.

Guillaume Gravier is a member of the Board of the technology cluster Images & Réseaux.

Guillaume Gravier is a member of the Board of the Comité des Projets of Inria - Rennes Bretagne Atlantique.

Pascale Sébillot is a member of the Conseil National des Universités 27th section (computer science).

Pascale Sébillot is the director of the Computer Science Laboratory, INSA Rennes.

Pascale Sébillot is the deputy director of the Scientific Advisory Committee of IRISA UMR 6074. Pascale Sébillot is a member of the theses advisory committee of the MathSTIC doctoral school. Pascale Sébillot is a member of the board of the MathSTIC doctoral school.

### **10.2. Teaching - Supervision - Juries**

### 10.2.1. Teaching

For reseachers, all activities are given. For professors and assistant professors, only courses at the M. Sc. level are listed.

Licence: Teddy Furon, Probabilities, 40h, L1, Agrocampus Rennes, France Licence: Guillaume Gravier, Databases, 30h, L2, INSA Rennes, France Licence: Guillaume Gravier, Probability and statistics, 16h, L3, INSA Rennes, France Licence: Guillaume Gravier, Natural Language Processing, 12h, L3 & M1, INSA Rennes, France Master: Laurent Amsaleg, Multidimensional indexing, 13h, M2, University Rennes 1, France Master: Yannis Avrithis, Deep Learning for Vision, 30h, M2, Univ. Rennes 1, France Master: Vincent Claveau, Data-Based Knowledge Acquisition: Symbolic Methods, 20h, L3, INSA Rennes, France Master: Vincent Claveau, Text Mining, 18h, M2, Univ. Rennes 1, France Master: Vincent Claveau, Information Retrieval, 15h, M2, ENSSAT, France Master: Vincent Claveau, Information Retrieval, 13h, M2, Univ. Rennes 1, France Master: Teddy Furon, Rare events, 20h, M2, Insa Rennes, France Master: Guillaume Gravier, Data analysis and probabilistic modeling, 30h, M2, University Rennes 1, France Master: Ewa Kijak, Image processing, 67h, M1, ESIR, France Master: Ewa Kijak, Supervised machine learning, 15h, M2R, University Rennes 1, France Master: Ewa Kijak, Supervised machine learning, 45h, M1, ESIR, France Master: Ewa Kijak, Image indexing, 17h, M2, University Rennes 1, France Master: Ewa Kijak, Indexing and multimedia databases, 15h, M2, ENSSAT, France Master: Ewa Kijak, Computer vision, 22h, M2, ESIR, France Master: Ewa Kijak, Image and text mining, 12h, M2, ENSAI, France Master: Simon Malinowski, Short-term time series prediction, 29h, M1, Univ. Rennes 1 Master: Simon Malinowski, Supervised Learning, 24h, M2, Univ. Rennes 1 Master: Christian Raymond, Dialogue, 4H, M2, University Rennes 1, France Master: Christian Raymond, Dialogue, 5H, M1, INSA Rennes, France Master: Pascale Sébillot, Advanced Databases and Modern Information Systems, 70h, M2, INSA Rennes, France Master: Pascale Sébillot, Logic Programming, 12h, M1, INSA Rennes, France Master: Pascale Sébillot, Natural Language Programming, 6h, M1, INSA Rennes, France

### 10.2.2. Supervision

PhD: Rémi Bois, Navigable directed multimedia hypergraphs: construction and exploitation, defended December 2017, Guillaume Gravier and Pascale Sébillot

PhD: Ahmet Iscen, Continuous memories for representing sets of vectors and image collections, defended September 2017, Teddy Furon

PhD: Raheel Kareem Qader, Phonology modeling for emotional speech synthesis, defended March 2017, Gwénolé Lecorvé and Pascale Sébillot (with EXPRESSION, IRISA team)

PhD: Vedran Vukotić, Deep neural architectures for automatic representation learning from multimedia multimodal data, defended September 2017, Guillaume Gravier and Christian Raymond

PhD in progress: Ricardo Carlini Sperandio, Unsupervised motif mining in multimedia time series, started August 2015, Laurent Amsaleg and Guillaume Gravier

PhD in progress: Clément Dalloux, Clinical text mining and indexing, started Dec. 2016, Vincent Claveau, Natalia Grabar (STL, Lille), Olivier Dameron (with DYLISS project-team)

PhD in progress: Mikail Demirdelen, User-adapted multi-document multimedia synthesis, started Oct. 2016, Guillaume Gravier and Pascale Sébillot

PhD in progress: Cheikh Brahim El Vaigh, Incremental content to data linking leveraging ontological knowledge in data journalism, started Oct. 2017, Guillaume Gravier, Pascale Sébillot and François Goasdoué (with CEDAR, Inria team)

PhD in progress: Mathieu Laroze, Active learning on adaptive representations for object detection in high-resolution imaging, started June 2016, Romain Dambreville, Chloe Friguet, Ewa Kijak and Sebastien Lefevre (with OBELIX, IRISA team)

PhD in progress: Cédric Maigrot, Detecting fake information on social networks, started October 2015, Laurent Amsaleg, Vincent Claveau and Ewa Kijak

PhD in progress: Antoine Perquin, Universal speech synthesis through embeddings of massive heterogeneous data, started Oct. 2017, Laurent Amsaleg, Gwénolé Lecorvé, Damien Lolive (with EXPRESSION, IRISA team)

PhD in progress: Oriane Simeoni, Invariance and supervision in visual learning, started Oct. 2016, Yannis Avrithis and Guillaume Gravier

PhD in progress: Hanwei Zhang, Deep learning in adversarial contexts, started Oct. 2017, Laurent Amsaleg, Teddy Furon, Ewa Kijak

### 10.2.3. Juries

Guillaume Gravier

PhD, reviewer, Aparna Nurani Venkitasubramanian, Katholieke Universiteit Leuven

PhD, reviewer, Heider Sanchez, Universidad de Chile

PhD, reviewer, Mateusz Budnik, Univ. Grenoble Alpes

PhD, reviewer, Nhi Tran, CNAM

PhD, president, Gaël Le Lan, Univ. du Maine

#### Ewa Kijak

PhD, member, Yanwei Cui, Université Bretagne Sud

#### Pascale Sébillot

PhD, reviewer, Dialekti Valsamou, Université Paris Sud

PhD, reviewer, Anaïs Ollagnier, Aix-Marseille Université

PhD, president, Joseph Lark, Université de Nantes

PhD, member, Yoann Dupont, Université Sorbonne Nouvelle - Paris 3

### **10.3.** Popularization

Vincent Claveau

participated to the collective book "Les Big Data à découvert", CNRS éditions got an interview published in CNRS - Le journal about detecting fake news and image forgeries (joint work with C. Maigrot and E. Kijak).

### **11. Bibliography**

### **Publications of the year**

### **Doctoral Dissertations and Habilitation Theses**

- [1] A. ISCEN. Continuous memories for representing sets of vectors and image collections, Université Rennes 1, September 2017, https://tel.archives-ouvertes.fr/tel-01661318.
- [2] A. ISCEN. Continuous memories for representing sets of vectors and image collections, Université Rennes 1, September 2017, https://tel.archives-ouvertes.fr/tel-01661319.
- [3] V. VUKOTIC. Deep neural architectures for automatic representation learning from multimedia multimodal data, INSA Rennes, September 2017, https://hal.inria.fr/tel-01629669.
- [4] V. VUKOTIC. Deep Neural Architectures for Automatic Representation Learning from Multimedia Multimodal Data, INSA de Rennes, September 2017, https://tel.archives-ouvertes.fr/tel-01662594.

### **Articles in International Peer-Reviewed Journal**

- [5] L. AMSALEG, V. CLAVEAU.Notes about the Computational Journalism workshop 2017, in "Recherche d'Information, Document et Web Sémantique", August 2017, vol. 1, n<sup>o</sup> 1, 6, https://hal.archives-ouvertes. fr/hal-01643444.
- [6] A. ISCEN, T. FURON, V. GRIPON, M. RABBAT, H. JÉGOU.Memory vectors for similarity search in highdimensional spaces, in "IEEE Transactions on Big Data", 2017 [DOI: 10.1109/TBDATA.2017.2677964], https://hal.inria.fr/hal-01481220.
- [7] M. LARSON, M. SOLEYMANI, G. GRAVIER, B. IONESCU, G. J. F. JONES. *The Benchmarking Initiative for Multimedia Evaluation: MediaEval 2016*, in "IEEE MultiMedia", January 2017, vol. 24, n<sup>o</sup> 1, p. 93 96 [DOI: 10.1109/MMUL.2017.9], https://hal.archives-ouvertes.fr/hal-01635965.
- [8] C. MAIGROT, S. BRINGAY, J. AZÉ. Concept drift vs suicide: How one can help prevent the other?, in "International Journal of Computational Linguistics and Applications", October 2017, vol. 8, n<sup>o</sup> 1, https:// hal.archives-ouvertes.fr/hal-01617870.
- [9] B. QU, F. VALLET, J. CARRIVE, G. GRAVIER. Content-based unsupervised segmentation of recurrent TV programs using grammatical inference, in "Multimedia Tools and Applications", 2017 [DOI: 10.1007/s11042-017-4816-5], https://hal.archives-ouvertes.fr/hal-01522688.
- [10] G. SARGENT, F. BIMBOT, E. VINCENT. Estimating the structural segmentation of popular music pieces under regularity constraints, in "IEEE/ACM Transactions on Audio, Speech, and Language Processing", 2017, https://hal.inria.fr/hal-01403210.

### **Articles in Non Peer-Reviewed Journal**

[11] Y. KOMPATSIARIS, G. GRAVIER. Special Issue on Content Based Multimedia Indexing, in "Multimedia Tools and Applications", 2017 [DOI: 10.1007/s11042-017-5106-Y], https://hal.archives-ouvertes.fr/hal-01578925.

### **International Conferences with Proceedings**

[12] Best Paper

O. AHMED, G. SARGENT, F. GARNIER, B. HUET, V. CLAVEAU, L. COUTURIER, R. TRONCY, G. GRAVIER, P. BOUZY, F. LEMÉNOREL.*NexGenTV: Providing Real-Time Insight during Political Debates in a Second Screen Application*, in "MM 2017 - 25th ACM International Conference on Multimedia", Moutain View, United States, October 2017, https://hal.archives-ouvertes.fr/hal-01635966.

- [13] L. AMSALEG, J. E. BAILEY, D. BARBE, S. ERFANI, M. E. HOULE, V. NGUYEN, M. RADOVANOVIC.*The Vulnerability of Learning to Adversarial Perturbation Increases with Intrinsic Dimensionality*, in "WIFS 2017 9th IEEE International Workshop on Information Forensics and Security", Rennes, France, Proceedings of the 9th IEEE Workshop on Information Forensics and Security, December 2017, https://hal.archives-ouvertes. fr/hal-01599355.
- [14] A.-M. A. AWAL, N. GHANMI, R. SICRE, T. FURON. Complex Document Classification and Localization Application on Identity Document Images, in "ICDAR 2017 - The 14th IAPR International Conference on Document Analysis and Recognition", Kyoto, Japan, November 2017, p. 1-6, https://hal.inria.fr/hal-01660504.
- [15] R. BOIS, G. GRAVIER, E. JAMET, E. MORIN, M. ROBERT, P. SÉBILLOT. Linking multimedia content for efficient news browsing, in "ICMR'17 - ACM International Conference on Multimedia Retrieval", Bucharest, Romania, June 2017, https://hal.archives-ouvertes.fr/hal-01522413.
- [16] R. BOIS, G. GRAVIER, E. JAMET, M. ROBERT, E. MORIN, P. SÉBILLOT. Language-based Construction of Explorable News Graphs for Journalists, in "Empirical Methods in Natural Language Processing - Workshop on Natural Language Processing meets Journalism", Copenhague, Denmark, 2017, https://hal.archivesouvertes.fr/hal-01578932.
- [17] R. BOIS, V. VUKOTIC, A.-R. SIMON, R. SICRE, C. RAYMOND, P. SÉBILLOT, G. GRAVIER. *Exploiting Multimodality in Video Hyperlinking to Improve Target Diversity*, in "MMM2017 International Conference on Multimedia Modeling", Reykyavik, Iceland, L. AMSALEG, G. P. GUðMUNDSSON, C. GURRIN, S. SATOH (editors), Lecture Notes in Computer Science, Springer, January 2017, vol. 10133 [DOI : 10.1007/978-3-319-51814-5\_16], https://hal.archives-ouvertes.fr/hal-01498130.
- [18] V. CLAVEAU. Clustering multi-relationnal TV data by diverting supervised ILP, in "ILP 2017 27th International Conference on Inductive Logic Programming", Orléans, France, Proceedings of the 27th International Conference on Inductive Logic Programming, September 2017, p. 1-6, https://hal.archives-ouvertes.fr/hal-01643977.
- [19] V. CLAVEAU, E. KIJAK. Strategies to select examples for Active Learning with Conditional Random Fields, in "CICLing 2017 - 18th International Conference on Computational Linguistics and Intelligent Text Processing", Budapest, Hungary, April 2017, p. 1-14, https://hal.archives-ouvertes.fr/hal-01621338.

- [20] V. CLAVEAU, L. E. S. OLIVEIRA, G. BOUZILLÉ, M. CUGGIA, C. M. CABRAL MORO, N. GRABAR.*Numerical eligibility criteria in clinical protocols: annotation, automatic detection and interpretation*, in "AIME 2017 16th Conference in Artificial Intelligence in Medecine", Vienne, Austria, LNCS, Springer, June 2017, vol. 10259, p. 203-208 [DOI: 10.1007/978-3-319-59758-4\_22], https://hal.archives-ouvertes.fr/hal-01618930.
- [21] M. DINARELLI, V. VUKOTIC, C. RAYMOND.Label-dependency coding in Simple Recurrent Networks for Spoken Language Understanding, in "Interspeech", Stockholm, Sweden, August 2017, https://hal.inria.fr/hal-01553830.
- [22] G. B. D. FONSECA, G. SARGENT, I. L. FREIRE, R. SICRE, Z. J. PATROCINIO, G. GRAVIER, S. J. F. GUIMARÃES. *Tag Propagation Approaches within Speaking Face Graphs for Multimodal Person Discovery*, in "International workshop on Content-Based Multimedia Indexing (CBMI)", Firenze, Italy, Proceedings of the 15th international workshop on Content-Based Multimedia Indexing, June 2017, https://hal.archives-ouvertes. fr/hal-01551648.
- [23] T. FURON. About zero bit watermarking error exponents, in "ICASSP2017 IEEE International Conference on Acoustics, Speech and Signal Processing", New Orleans, United States, IEEE, March 2017, https://hal.inria. fr/hal-01512705.
- [24] G. Þ. GUðMUNDSSON, L. AMSALEG, B. Þ. JÓNSSON, M. J. FRANKLIN. Towards Engineering aWeb-Scale Multimedia Service: A Case Study Using Spark, in "Multimedia Systems Conference", Taipei, Taiwan, MMSys'17 Proceedings of the 8th ACM on Multimedia Systems Conference, acm, June 2017 [DOI: 10.1145/3083187.3083200], https://hal.archives-ouvertes.fr/hal-01565196.
- [25] A. ISCEN, G. TOLIAS, Y. AVRITHIS, T. FURON, O. CHUM. Efficient Diffusion on Region Manifolds: Recovering Small Objects with Compact CNN Representations, in "2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)", Honolulu, United States, July 2017, https://hal.inria.fr/hal-01505470.
- [26] A. ISCEN, G. TOLIAS, Y. AVRITHIS, T. FURON, O. CHUM. Panorama to panorama matching for location recognition, in "ACM International Conference on Multimedia Retrieval (ICMR) 2017", Bucharest, Romania, June 2017 [DOI: 10.1145/3078971.3079033], https://hal.inria.fr/hal-01511793.
- [27] N. LE, H. BREDIN, G. SARGENT, M. INDIA, P. LOPEZ-OTERO, C. BARRAS, C. GUINAUDEAU, G. GRAVIER, G. BARBOSA DA FONSECA, I. LYON FREIRE, Z. PATROCÍNIO JR, S. J. F. GUIMARÃES, G. MARTÍ, J. R. MORROS, J. HERNANDO, L. DOCIO-FERNANDEZ, C. GARCIA-MATEO, S. MEIGNIER, J.-M. ODOBEZ. *Towards large scale multimedia indexing: A case study on person discovery in broadcast news*, in "Content-Based Multimedia Indexing CBMI", Firenze, Italy, Proceedings of the 15th international workshop on Content-Based Multimedia Indexing, June 2017 [*DOI :* 10.1145/3095713.3095732], https://hal.archives-ouvertes.fr/hal-01551690.
- [28] Best Paper

A. LODS, S. MALINOWSKI, R. TAVENARD, L. AMSALEG.*Learning DTW-Preserving Shapelets*, in "IDA 2017 - 16th International Symposium on Intelligent Data Analysis", London, United Kingdom, Advances in Intelligent Data Analysis XVI, springer International Publishing, October 2017, vol. 10584, p. 198-209 [*DOI* : 10.1007/978-3-319-68765-0\_17], https://hal.archives-ouvertes.fr/hal-01565207.

- [29] C. MAIGROT, E. KIJAK, R. SICRE, V. CLAVEAU. *Tampering detection and localization in images from social networks: A CBIR approach*, in "ICIAP 2017 International Conference on Image Analysis and Processing", Catane, Italy, September 2017, p. 1-11, https://hal.inria.fr/hal-01623105.
- [30] H. PÁLMASSON, B. Þ. JÓNSSON, L. AMSALEG, M. SCHEDL, P. KNEES. On Competitiveness of Nearest-Neighbor Based Music Classification: A Methodological Critique, in "10th International Conference on Similarity Search and Applications", munich, Germany, 10th International Conference on Similarity Search and Applications, October 2017, https://hal.archives-ouvertes.fr/hal-01565217.
- [31] R. QADER, G. LECORVÉ, D. LOLIVE, M. TAHON, P. SÉBILLOT. Statistical Pronunciation Adaptation for Spontaneous Speech Synthesis, in "Text, Speech and Dialogue (TSD)", Prague, Czech Republic, August 2017, https://hal.inria.fr/hal-01532035.
- [32] R. SICRE, Y. AVRITHIS, E. KIJAK, F. JURIE. Unsupervised part learning for visual recognition, in "CVPR 2017 - IEEE Conference on Computer Vision and Pattern Recognition", Honolulu, United States, July 2017, https://arxiv.org/abs/1704.03755, https://hal.archives-ouvertes.fr/hal-01507379.
- [33] R. TAVENARD, S. MALINOWSKI, L. CHAPEL, A. BAILLY, H. SANCHEZ, B. BUSTOS. Efficient Temporal Kernels between Feature Sets for Time Series Classification, in "European Conference on Machine Learning and Principles and Practice of Knowledge Discovery", Skopje, Macedonia, September 2017, https://halshs. archives-ouvertes.fr/halshs-01561461.
- [34] V. VUKOTIC, S.-L. PINTEA, C. RAYMOND, G. GRAVIER, J. C. VAN GEMERT. One-Step Time-Dependent Future Video Frame Prediction with a Convolutional Encoder-Decoder Neural Network, in "International Conference of Image Analysis and Processing (ICIAP)", Catania, Italy, Proceedings of the 19th International Conference of Image Analysis and Processing, September 2017, https://hal.inria.fr/hal-01567742.
- [35] V. VUKOTIC, C. RAYMOND, G. GRAVIER. Generative Adversarial Networks for Multimodal Representation Learning in Video Hyperlinking, in "ACM International Conference on Multimedia Retrieval (ICMR) 2017", Bucharest, Romania, Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval, ACM, June 2017 [DOI: 10.1145/3078971.3079038], https://hal.inria.fr/hal-01522419.

### **National Conferences with Proceeding**

- [36] V. CLAVEAU, C. RAYMOND.IRISA at DeFT2017 : classification systems of increasing complexity, in "DeFT 2017 Défi Fouille de texte", Orléans, France, Actes de l'Atelier Défi Fouille de Texte, DeFT, June 2017, p. 1-10, https://hal.archives-ouvertes.fr/hal-01643993.
- [37] C. DALLOUX, V. CLAVEAU, N. GRABAR. Détection de la négation : corpus français et apprentissage supervisé, in "SIIM 2017 - Symposium sur l'Ingénierie de l'Information Médicale", Toulouse, France, November 2017, p. 1-8, https://hal.archives-ouvertes.fr/hal-01659637.
- [38] C. MAIGROT, V. CLAVEAU, E. KIJAK. Détection de fausses informations dans les réseaux sociaux : vers des approches multi-modales, in "Extraction et Gestion des Connaissances (EGC)", Grenoble, France, January 2017, https://hal.inria.fr/hal-01548156.
- [39] C. MAIGROT, E. KIJAK, V. CLAVEAU. Détection de fausses informations dans les réseaux sociaux : l'utilité des fusions de connaissances, in "COnférence Recherche d'Information et Applications", Marseille, France,

Actes de la conférence CORIA 2017, March 2017, p. 107-122 [*DOI* : 10.24348/CORIA.2017.11], https://hal.archives-ouvertes.fr/hal-01617878.

### [40] Best Paper

R. QADER, G. LECORVÉ, D. LOLIVE, P. SÉBILLOT. *Automatic disfluency insertion towards spontaneous TTS : formalization and proof of concept*, in "Traitement automatique du langage naturel (TALN)", Orléans, France, June 2017, https://hal.inria.fr/hal-01532031.

### **Conferences without Proceedings**

- [41] C. DALLOUX.*Identifying uncertainty and negation's cues and scope : State of the art*, in "RECITAL 2017 18ème Rencontre des Étudiants Chercheurs en Informatique en Traitement Automatique des Langues", Orléans, France, June 2017, p. 1-14, https://hal.archives-ouvertes.fr/hal-01659646.
- [42] M. DEMIRDELEN, M. BUDNIK, G. SARGENT, R. BOIS, G. GRAVIER. IRISA at TRECVid 2017: Beyond Crossmodal and Multimodal Models for Video Hyperlinking, in "Working Notes of the TRECVid 2017 Workshop", Gettysburg, United States, 2017, https://hal.archives-ouvertes.fr/hal-01643232.
- [43] N. GRABAR, V. CLAVEAU. Numerical criteria in clinical trials: annotating, detecting and normalizing, in "Actes de la conférence TALN 2017", Orléans, France, June 2017, https://hal.archives-ouvertes.fr/hal-01643969.
- [44] R. SICRE, A. MONTASER AWAL, T. FURON.*Identity documents classification as an image classification problem*, in "ICIAP 2017 19th International Conference on Image Analysis and Processing", Catania, Italy, ICIAP 2017: Image Analysis and Processing ICIAP 2017, Springer, September 2017, p. 602-613 [DOI: 10.1007/978-3-319-68548-9\_55], https://hal.archives-ouvertes.fr/hal-01660500.
- [45] R. SICRE, J. J. RABIN, Y. AVRITHIS, T. FURON, F. JURIE, E. KIJAK. Automatic discovery of discriminative parts as a quadratic assignment problem, in "ICCV Workshops – CEFRL", Venise, Italy, October 2017, https:// hal.archives-ouvertes.fr/hal-01623148.

### **Books or Proceedings Editing**

- [46] L. AMSALEG, V. CLAVEAU, X. TANNIER (editors). *Proceedings of the Computational Journalism workshop* 2017, January 2017, https://hal.archives-ouvertes.fr/hal-01643634.
- [47] L. AMSALEG, G. Þ. GUÖMUNDSSON, C. GURRIN, B. Þ. JÓNSSON, S. SATOH (editors). Proceedings of the 23rd International Conference on Multimedia Modeling, MMM 2017, Reykjavik, Iceland, January 4-6, 2017, MultiMedia Modeling, Springer International Publishing, Reykjavík, Iceland, January 2017, vol. 10132, n<sup>o</sup> 0302-9743 [DOI: 10.1007/978-3-319-51811-4], https://hal.archives-ouvertes.fr/hal-01430555.
- [48] G. GRAVIER, B. BISHKE, C.-H. DEMARTY, M. ZAHARIEVA, M. A. RIEGLER, E. DELLANDRÉA, D. BOGDANOV, S. RICHARD, G. J. F. JONES, M. LARSON (editors). Working Notes Proceedings of the MediaEval Workshop, November 2017, https://hal.archives-ouvertes.fr/hal-01643258.

### **Research Reports**

[49] R. SICRE, A. M. AWAL, T. FURON. Identity documents classification as an image classification problem, Inria Rennes - Bretagne Atlantique; IRISA; AriadNext, April 2017, n<sup>o</sup> RT-0488, https://hal.inria.fr/hal-01503541.

### **References in notes**

- [50] M. AGOSTI, J. ALLAN. Special issue on methods and tools for the automatic construction of hypertext, in "Information Processing and Retrieval", 1997, vol. 33, n<sup>o</sup> 2.
- [51] C. A. BHATT, M. S. KANKANHALLI. Multimedia data mining: State of the art and challenges, in "Multimedia Tools and Applications", 2011, vol. 51, n<sup>o</sup> 1, p. 35–76.
- [52] O. CHUM, J. MATAS.Large-Scale Discovery of Spatially Related Images, in "Pattern Analysis and Machine Intelligence", 2010, vol. 32, n<sup>o</sup> 2, p. 371-377.
- [53] M. ESKEVICH, G. J. F. JONES, S. CHEN, R. ALY, R. ORDELMAN, D. NADEEM, C. GUINAUDEAU, G. GRAVIER, P. SÉBILLOT, T. DE NIES, P. DEBEVERE, R. V. DE WALLE, P. GALUŠCKOVÁ, P. PECINA, M. LARSON. *Multimedia Information Seeking through Search And Hyperlinking*, in "Intl. Conf. on Multimedia Retrieval", 2013.
- [54] C. HERLEY.ARGOS: Automatically extracting Repeating Objects from multimedia Streams, in "IEEE Trans. on Multimedia", Feb. 2006, vol. 8, n<sup>o</sup> 1, p. 115–129.
- [55] A. PARK, J. R. GLASS. Unsupervised pattern discovery in speech, in "IEEE Transaction on Acoustic, Speech and Language Processing", Jan. 2008, vol. 16, n<sup>o</sup> 1, p. 186–197.
- [56] K. M. PUA, J. GAUCH, S. E. GUCH, J. Z. MIADOWICZ. Real time repeated video sequence identification, in "Computer Vision and Image Understanding", 2004, vol. 93, n<sup>o</sup> 2004, p. 310–327.
- [57] C. G. M. SNOEK, M. WORRING. *Multimodal Video Indexing: A Review of the State-of-the-Art*, in "Multimedia Tools and Applications", 2005, vol. 25, n<sup>o</sup> 1, p. 5–35.
- [58] O. DE ROOIJ, M. WORRING. Browsing Video Along Multiple Threads, in "IEEE Transactions on Multimedia", 2010, vol. 12, n<sup>o</sup> 2, p. 121–130.

# **Project-Team MIMETIC**

## Analysis-Synthesis Approach for Virtual Human Simulation

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: Université Haute Bretagne (Rennes 2) Université Rennes 1 École normale supérieure de Rennes

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Interaction and visualization
# **Table of contents**

1.	Personnel			
2.	Overall Objectives			
3.	Research Program			
	3.1. Biomechanics and Motion Control	654		
	3.2. Experiments in Virtual Reality	656		
	3.3. Computational Geometry	656		
4.	Application Domains	657		
	4.1. Autonomous Characters	657		
	4.2. Biomechanics and Motion Analysis	657		
	4.3. Interactions between walkers	658		
	4.4. Motion Sensing of Human Activity	658		
	4.5. VR and Sports	659		
	4.6. Interactive Digital Storytelling	659		
	4.7. VR and Ergonomics	660		
5.	Highlights of the Year	660		
6.	New Software and Platforms	661		
	6.1. AsymGait	661		
	6.2. Cinematic Viewpoint Generator	661		
	6.3. Directors Lens Motion Builder	661		
	6.4. Kimea	662		
	6.5. Populate	662		
	6.6. The Theater	663		
	6.7. CusToM	663		
	6.8. MotionGraphVR	663		
	6.9. Platforms	664		
7.	New Results	664		
	7.1. Outline	664		
	7.2. Motion analysis	664		
	7.2.1. Biomechanics for motion analysis-synthesis	664		
	7.2.2. Interactions between walkers	665		
	7.2.3. New automatic methods to assess motion in industrial contexts based on Kinect	666		
	7.2.4. Clinical gait assessment based on Kinect data	667		
	7.2.5. Biomechanical analysis of tennis serve	668		
	7.3. Virtual human simulation			
	7.4. Human motion in VR	669		
	7.4.1. Motion recognition and classification	669		
	7.4.2. Automatic evaluation of sports gesture	669		
	7.4.3. Biofidelity in VR	669		
	7.5. Digital story telling	669		
	7.5.1. Analysis of Film Style	669		
	7.5.2. Film Editing Patterns: Thinking like a Director	670		
	7.5.3. Directing Cinematographic Drones	670		
8.	Bilateral Contracts and Grants with Industry	670		
	8.1. Bilateral Contracts with Industry	670		
	8.2. Bilateral Grants with Industry   67			
9.	Partnerships and Cooperations			
	9.1. Regional Initiatives			
	9.2. National Initiatives	671		
	9.2.1. ANR	671		

	9.2.1.1.	Cineviz	671
	9.2.1.2.	Entracte	672
	9.2.2. Natio	onal scientific collaborations	672
	9.2.2.1.	Cavaletic	672
	9.2.2.2.	FFT	672
	9.2.2.3.	gDGA	672
	9.2.2.4.	IRMA	672
	9.2.3. ADT	': Immerstar	673
	9.2.4. PRE		673
	9.3. Internatio	nal Initiatives	673
	9.3.1. FOR	RMOSA	673
	9.3.2. SIM	S	674
	9.3.3. Inria	International Partners	674
	9.4. Internatio	nal Research Visitors	674
	9.4.1.1.	Professors and associate professors	674
	9.4.1.2.	Internships	674
10.	Dissemination		
10.1. Promoting Scientific Activities			674
	10.1.1. Scier	tific Events Organisation	674
	10.1.1.1.	General Chair, Scientific Chair	674
	10.1.1.2.	Member of the Organizing Committees	675
10.1.2. Scientific Events Selection			675
	10.1.2.1.	Chair of Conference Program Committees	675
	10.1.2.2.	Member of the Conference Program Committees	675
	10.1.2.3.	Reviewer	675
	10.1.3. Journ	nal	676
	10.1.3.1.	Member of the Editorial Boards	676
	10.1.3.2.	Reviewer - Reviewing Activities	676
	10.1.4. Invite	ed Talks	676
	10.1.5. Scier	tific Expertise	676
	10.1.6. Rese	arch Administration	676
	10.2. Teaching	- Supervision - Juries	677
	10.2.1. Teach	ning	677
10.2.2. Supervision			679
	10.2. Derevlarization		
11	10.3. Populariz	ation	681
11.	Bibliography		

# **Project-Team MIMETIC**

*Creation of the Team: 2011 January 01, updated into Project-Team: 2014 January 01* **Keywords:** 

### **Computer Science and Digital Science:**

A5.1.5. - Body-based interfaces
A5.4.2. - Activity recognition
A5.4.5. - Object tracking and motion analysis
A5.5.4. - Animation
A5.6. - Virtual reality, augmented reality
A5.10.3. - Planning
A5.10.5. - Robot interaction (with the environment, humans, other robots)
A5.11.1. - Human activity analysis and recognition
A6. - Modeling, simulation and control

Other Research Topics and Application Domains:

- B1.2.2. Cognitive science
- B2.5. Handicap and personal assistances
- B2.8. Sports, performance, motor skills
- B5.1. Factory of the future
- B5.8. Learning and training
- B7.1.1. Pedestrian traffic and crowds
- B9.2.2. Cinema, Television
- B9.2.3. Video games
- B9.3. Sports

# 1. Personnel

### **Research Scientist**

Ludovic Hoyet [Inria, Researcher]

### **Faculty Members**

Franck Multon [Team leader, Université Rennes 2, Professor, HDR] Benoit Bideau [Université Rennes 2, Professor, HDR] Nicolas Bideau [Université Rennes 2, Associate Professor] Marc Christie [Université de Rennes 1, Associate Professor] Armel Crétual [Université Rennes 2, Associate Professor, HDR] Georges Dumont [Ecole normale supérieure de Rennes, Professor, HDR] Richard Kulpa [Université Rennes 2, Associate Professor, HDR] Fabrice Lamarche [Université de Rennes 1, Associate Professor] Antonio Mucherino [Université de Rennes 1, Associate Professor] Guillaume Nicolas [Université Rennes 2, Associate Professor] Anne-Hélène Olivier [Université Rennes 2, Associate Professor] Anne-Hélène Olivier [Université Rennes 2, Associate Professor] Anne-Hélène Olivier [Université Rennes 2, Associate Professor] Antonine Muller [Ecole normale supérieure de Rennes, Associate Professor] Antoine Muller [Ecole normale supérieure de Rennes, Associate Professor] Antoine Muller [Ecole normale supérieure de Rennes, Associate Professor] Antoine Muller [Ecole normale supérieure de Rennes, Associate Professor] Antoine Muller [Ecole normale supérieure de Rennes, Associate Professor] Antoine Muller [Ecole normale supérieure de Rennes, Associate Professor] Hui-Yin Wu [Université de Rennes 1, Assistant Professor, until September 2017]

#### **Technical Staff**

Ronan Gaugne [Université de Rennes 1, SED Research Engeneer, involved in MimeTIC at 15%] Anthony Sorel [Université de Rennes 2, Engeneer] Quentin Galvane [Inria] Florian Brillat [Inria, Ecole normale supérieure de Rennes, from October 2017] Christophe Lino [Inria, Engeneer, until Sep 2017]

#### **PhD Students**

Amaury Louarn [Université de Rennes 1, from October 2017] Marion Morel [Université de Rennes 2, until November 2017] Yacine Boulahia [INSA Rennes] Florian Berton [Inria, from November 2017] Rebecca Fribourg [Inria, from September 2017] Charles Faure [Ecole normale supérieure de Rennes] Simon Hilt [Ecole normale supérieure de Rennes, from Sep 2017] Sean Lynch [Univ de Haute-Bretagne] Theo Perrin [Ecole normale supérieure de Rennes, from Sep 2017] Pierre Puchaud [Saint Cyr, from Oct 2017]

#### **Post-Doctoral Fellows**

Diane Haering [Inria, until November 2017] Laurentius Meerhoff [Inria, until July 2017] Xi Wang [Inria]

#### Visiting Scientists

I-Sheng Lin [National Cheng Chi University, from Jul 2017 until Sep 2017] Emma Carrigan [Trinity College, from January to June 2017]

#### Administrative Assistant

Nathalie Denis [Inria, Administrative Assistant]

# 2. Overall Objectives

# 2.1. Presentation

MimeTIC is a multidisciplinary team whose aim is to better understand and model human activity in order to simulate realistic autonomous virtual humans: realistic behaviors, realistic motions and realistic interactions with other characters and users. It leads to modeling the complexity of a human body, as well as of his environment where he can pick-up information and he can act on it. A specific focus is dedicated to human physical activity and sports as it raises the highest constraints and the highest complexity when addressing these problems. Thus, MimeTIC is composed of experts in computer science whose research interests are computer animation, behavioral simulation, motion simulation, crowds and interaction between real and virtual humans. MimeTIC is also composed of experts in sports science, motion analysis, motion sensing, biomechanics and motion control. Hence, the scientific foundations of MimeTIC are motion sciences (biomechanics, motion control, perception-action coupling, motion analysis), computational geometry (modeling of the 3D environment, motion planning, path planning) and design of protocols in immersive environments (use of virtual reality facilities to analyze human activity).

Thanks to these skills, we wish to reach the following objectives: to make virtual humans behave, move and interact in a natural manner in order to increase immersion and to improve knowledge on human motion control. In real situations (see Figure 1), people have to deal with their physiological, biomechanical and neurophysiological capabilities in order to reach a complex goal. Hence MimeTIC addresses the problem of modeling the anatomical, biomechanical and physiological properties of human beings. Moreover these characters have to deal with their environment. Firstly they have to perceive this environment and pick-up relevant information. MimeTIC thus addresses the problem of modeling the environment including its geometry and associated semantic information. Secondly, they have to act on this environment to reach their goals. It leads to cognitive processes, motion planning, joint coordination and force production in order to act on this environment.



Figure 1. Main objective of MimeTIC: to better understand human activity in order to improve virtual human simulations. It involves modeling the complexity of human bodies, as well as of environments where to pick-up information and act upon.

In order to reach the above objectives, MimeTIC has to address three main challenges:

- dealing with the intrinsic complexity of human beings, especially when addressing the problem of
  interactions between people for which it is impossible to predict and model all the possible states of
  the system,
- making the different components of human activity control (such as the biomechanical and physical, the reactive, cognitive, rational and social layers) interact while each of them is modeled with completely different states and time sampling,
- and being able to measure human activity while balancing between ecological and controllable protocols, and to be able to extract relevant information in wide databases of information.

Contrary to many classical approaches in computer simulation, which mostly propose simulation without trying to understand how real people do, the team promotes a coupling between human activity analysis and synthesis, as shown in Figure 2.

In this research path, **improving knowledge on human activity** enables us to highlight fundamental assumptions about natural control of human activities. These contributions can be promoted in e.g. biomechanics, motion sciences, neurosciences. According to these assumptions we propose new algorithms for controlling **autonomous virtual humans**. The virtual humans can perceive their environment and decide of the most natural action to reach a given goal. This work is promoted in computer animation, virtual reality and has some applications in robotics through collaborations. Once autonomous virtual humans have the ability to act as real humans would in the same situation, it is possible to make them **interact with others**, i.e., autonomous characters (for crowds or group simulations) as well as real users. The key idea here is to analyze to what extent the assumptions proposed at the first stage lead to natural interactions with real users. This process enables the validation of both our assumptions and our models.



Figure 2. Research path of MimeTIC: coupling analysis and synthesis of human activity enables us to create more realistic autonomous characters and to evaluate assumptions about human motion control.

Among all the problems and challenges described above, MimeTIC focuses on the following domains of research:

- **motion sensing** which is a key issue to extract information from raw motion capture systems and thus to propose assumptions on how people control their activity,
- human activity & virtual reality, which is explored through sports application in MimeTIC. This domain enables the design of new methods for analyzing the perception-action coupling in human activity, and to validate whether the autonomous characters lead to natural interactions with users,
- **interactions** in small and large groups of individuals, to understand and model interactions with lot of individual variability such as in crowds,
- virtual storytelling which enables us to design and simulate complex scenarios involving several humans who have to satisfy numerous complex constraints (such as adapting to the real-time environment in order to play an imposed scenario), and to design the coupling with the camera scenario to provide the user with a real cinematographic experience,
- **biomechanics** which is essential to offer autonomous virtual humans who can react to physical constraints in order to reach high-level goals, such as maintaining balance in dynamic situations or selecting a natural motor behavior among the whole theoretical solution space for a given task,
- and **autonomous characters** which is a transversal domain that can reuse the results of all the other domains to make these heterogeneous assumptions and models provide the character with natural behaviors and autonomy.

# 3. Research Program

## **3.1. Biomechanics and Motion Control**

Human motion control is a highly complex phenomenon that involves several layered systems, as shown in Figure 3. Each layer of this controller is responsible for dealing with perceptual stimuli in order to decide the

actions that should be applied to the human body and his environment. Due to the intrinsic complexity of the information (internal representation of the body and mental state, external representation of the environment) used to perform this task, it is almost impossible to model all the possible states of the system. Even for simple problems, there generally exists an infinity of solutions. For example, from the biomechanical point of view, there are much more actuators (i.e. muscles) than degrees of freedom leading to an infinity of muscle activation patterns for a unique joint rotation. From the reactive point of view there exists an infinity of paths to avoid a given obstacle in navigation tasks. At each layer, the key problem is to understand how people select one solution among these infinite state spaces. Several scientific domains have addressed this problem with specific points of view, such as physiology, biomechanics, neurosciences and psychology.



Figure 3. Layers of the motion control natural system in humans.

In biomechanics and physiology, researchers have proposed hypotheses based on accurate joint modeling (to identify the real anatomical rotational axes), energy minimization, force and torques minimization, comfort maximization (i.e. avoiding joint limits), and physiological limitations in muscle force production. All these constraints have been used in optimal controllers to simulate natural motions. The main problem is thus to define how these constraints are composed altogether such as searching the weights used to linearly combine these criteria in order to generate a natural motion. Musculoskeletal models are stereotyped examples for which there exists an infinity of muscle activation patterns, especially when dealing with antagonist muscles. An unresolved problem is to define how to use the above criteria to retrieve the actual activation patterns, while optimization approaches still leads to unrealistic ones. It is still an open problem that will require multidisciplinary skills including computer simulation, constraint solving, biomechanics, optimal control, physiology and neuroscience.

In neuroscience, researchers have proposed other theories, such as coordination patterns between joints driven by simplifications of the variables used to control the motion. The key idea is to assume that instead of controlling all the degrees of freedom, people control higher level variables which correspond to combinations of joint angles. In walking, data reduction techniques such as Principal Component Analysis have shown that lower-limb joint angles are generally projected on a unique plane whose angle in the state space is associated with energy expenditure. Although knowledge exists for specific motions, such as locomotion or grasping, this type of approach is still difficult to generalize. The key problem is that many variables are coupled and it is very difficult to objectively study the behavior of a unique variable in various motor tasks. Computer simulation is a promising method to evaluate such type of assumptions as it enables to accurately control all the variables and to check if it leads to natural movements. Neuroscience also addresses the problem of coupling perception and action by providing control laws based on visual cues (or any other senses), such as determining how the optical flow is used to control direction in navigation tasks, while dealing with collision avoidance or interception. Coupling of the control variables is enhanced in this case as the state of the body is enriched by the large amount of external information that the subject can use. Virtual environments inhabited with autonomous characters whose behavior is driven by motion control assumptions is a promising approach to solve this problem. For example, an interesting problem in this field is navigation in an environment inhabited with other people. Typically, avoiding static obstacles together with other people displacing into the environment is a combinatory problem that strongly relies on the coupling between perception and action.

One of the main objectives of MimeTIC is to enhance knowledge on human motion control by developing innovative experiments based on computer simulation and immersive environments. To this end, designing experimental protocols is a key point and some of the researchers in MimeTIC have developed this skill in biomechanics and perception-action coupling. Associating these researchers to experts in virtual human simulation, computational geometry and constraints solving enable us to contribute to enhance fundamental knowledge in human motion control.

# **3.2. Experiments in Virtual Reality**

Understanding interactions between humans is challenging because it addresses many complex phenomena including perception, decision-making, cognition and social behaviors. Moreover, all these phenomena are difficult to isolate in real situations, and it is therefore highly complex to understand their individual influence on these human interactions. It is then necessary to find an alternative solution that can standardize the experiments and that allows the modification of only one parameter at a time. Video was first used since the displayed experiment is perfectly repeatable and cut-offs (stop the video at a specific time before its end) allow having temporal information. Nevertheless, the absence of adapted viewpoint and stereoscopic vision does not provide depth information that are very meaningful. Moreover, during video recording session, the real human is acting in front of a camera and not of an opponent. The interaction is then not a real interaction between humans.

Virtual Reality (VR) systems allow full standardization of the experimental situations and the complete control of the virtual environment. It is then possible to modify only one parameter at a time and to observe its influence on the perception of the immersed subject. VR can then be used to understand what information is picked up to make a decision. Moreover, cut-offs can also be used to obtain temporal information about when information is picked up. When the subject can moreover react as in a real situation, his movement (captured in real time) provides information about his reactions to the modified parameter. Not only is the perception studied, but the complete perception-action loop. Perception and action are indeed coupled and influence each other as suggested by Gibson in 1979.

Finally, VR allows the validation of virtual human models. Some models are indeed based on the interaction between the virtual character and the other humans, such as a walking model. In that case, there are two ways to validate it. First, they can be compared to real data (e.g. real trajectories of pedestrians). But such data are not always available and are difficult to get. The alternative solution is then to use VR. The validation of the realism of the model is then done by immersing a real subject in a virtual environment in which a virtual character is controlled by the model. Its evaluation is then deduced from how the immersed subject reacts when interacting with the model and how realistic he feels the virtual character is.

# 3.3. Computational Geometry

Computational geometry is a branch of computer science devoted to the study of algorithms which can be stated in terms of geometry. It aims at studying algorithms for combinatorial, topological and metric problems concerning sets of points in Euclidian spaces. Combinatorial computational geometry focuses on three main problem classes: static problems, geometric query problems and dynamic problems.

In static problems, some inputs are given and the corresponding outputs need to be constructed or found. Such problems include linear programming, Delaunay triangulations, and Euclidian shortest paths for instance. In geometric query problems, commonly known as geometric search problems, the input consists of two parts: the search space part and the query part, which varies over the problem instances. The search space typically needs to be preprocessed, in a way that multiple queries can be answered efficiently. Some typical problems are range searching, point location in a portioned space, or nearest neighbor queries. In dynamic problems, the goal is to find an efficient algorithm for finding a solution repeatedly after each incremental modification of the input data (addition, deletion or motion of input geometric elements). Algorithms for problems of this type typically involve dynamic data structures. Both of previous problem types can be converted into a dynamic problem, for instance, maintaining a Delaunay triangulation between moving points.

In this context, distance geometry relies solely on distances, instead of points and lines, as in classical geometry. Various applications lead to the definition of problems that can be formulated as a distance geometry, including sensor network localization, robot coordination, the identification of molecular conformations, or as in the context of MimeTIC relations between objects in virtual scenes (e.g., distances between body segments, agents, or cameras). In recent years, scientific research has been oriented to the assumptions allowing for discretizing the search space of a given distance geometry problem. The discretization (which is exact in some situations) allows to conceive ad-hoc and efficient algorithms, and for enumerating the entire solution set of a given instance.

The MimeTIC team works on problems such as crowd simulation, spatial analysis, path and motion planning in static and dynamic environments, camera planning with visibility constraints for instance. The core of those problems, by nature, relies on problems and techniques belonging to computational geometry. Proposed models pay attention to algorithms complexity to be compatible with performance constraints imposed by interactive applications.

# 4. Application Domains

### **4.1.** Autonomous Characters

Autonomous characters are becoming more and more popular as they are used in an increasing number of application domains. In the field of special effects, virtual characters are used to replace secondary actors and generate highly populated scenes that would be hard and costly to produce with real actors. In video games and virtual storytelling, autonomous characters play the role of actors that are driven by a scenario. Their autonomy allows them to react to unpredictable user interactions and adapt their behavior accordingly. In the field of simulation, autonomous characters are used to simulate the behavior of humans in different kind of situations. They enable to study new situations and their possible outcomes.

One of the main challenges in the field of autonomous characters is to provide a unified architecture for the modeling of their behavior. This architecture includes perception, action and decisional parts. This decisional part needs to mix different kinds of models, acting at different time scale and working with different nature of data, ranging from numerical (motion control, reactive behaviors) to symbolic (goal oriented behaviors, reasoning about actions and changes).

In the MimeTIC team, we focus on autonomous virtual humans. Our problem is not to reproduce the human intelligence but to propose an architecture making it possible to model credible behaviors of anthropomorphic virtual actors evolving/moving in real time in virtual worlds. The latter can represent particular situations studied by psychologists of the behavior or to correspond to an imaginary universe described by a scenario writer. The proposed architecture should mimic all the human intellectual and physical functions.

### 4.2. Biomechanics and Motion Analysis

Biomechanics is obviously a very large domain. This large set can be divided regarding to the scale at which the analysis is performed going from microscopic evaluation of biological tissues' mechanical properties to macroscopic analysis and modeling of whole body motion. Our topics in the domain of biomechanics mainly lie within this last scope. In order to obtain a better understanding of human motion, MimeTIC addresses three main situations: everyday motions of a lambda subject, locomotion of pathological subjects and sports gestures.

In the first situation, MimeTIC is interested in studying how subjects maintain their balance in highly dynamic conditions. Until now, balance have nearly always been considered in static or quasi-static conditions. The knowledge of much more dynamic cases still has to be improved. Our approach has demonstrated that, first of all, the question of the parameter that will allow to do this is still open. We have also largely contributed to gaining a better understanding of collision avoidance between pedestrians. This topic includes the research of the parameters that are interactively controlled and the study of each one's role within this interaction.

The second situation focuses on locomotion of pathological subjects. When patients cannot walk efficiently, in particular those suffering from central nervous system affections, it becomes very useful for practitioners to benefit from an objective evaluation of their capacities. To facilitate such evaluations, we have developed two complementary indices, one based on kinematics and the other one on muscle activations. One major point of our research is that such indices are usually only developed for children whereas adults with these affections are much more numerous. Finally, in sports, where gesture can be considered, in some way, as abnormal, the goal is more precisely to understand the determinants of performance. This could then be used to improve training programs or devices. Two different sports have been studied: a) the tennis serve, where the goal was to understand the contribution of each segment of the body on the speed of the ball and b) the influence of the mechanical characteristics of the fin in fin swimming.

After having improved the knowledge of these different gestures a second goal is then to propose modeling solutions that can be used in VR environments for other research topics within MimeTIC. This has been the case, for example, for collision avoidance.

### 4.3. Interactions between walkers

Modeling and simulating the interactions between walkers is a very active, complex and competitive domain, interesting various disciplines such as mathematics, cognitive sciences, physics, computer graphics, etc. Interactions between walkers are by definition at the very core of our society since they represent the basic synergies of our daily life. When walking in the street, we take information about our surrounding environment in order to interact with people, move without collision, alone or in a group, intercept, meet or escape to somebody. Large groups of walkers can be first seen as a complex system: numerous local interactions occur between its elements and result into macroscopic emergent phenomena. Interactions are of various nature (e.g., collision avoidance, following) and are undergoing various factors as well. Physical factors are crucial as a group gathers by definition numerous moving people with a certain level of density. But sociological, cultural and psychological factors are important as well, since people's behavior is deeply changed from country to country, or depending on the considered situations. On the computational point of view, simulating the movements of large groups of walkers (i.e., crowds) pushes traditional simulation algorithms to their limit. As an element of a crowd is subject to interact with any other element belonging the same crowd, a naïve simulation algorithm has a quadratic complexity. Specific strategies are set to face such a difficulty: level-of-detail techniques enable scaling large crowd simulation and reach real-time solutions.

MimeTIC is an international key contributor in the domain of understanding and simulating interactions between walkers, in particular for virtual crowds. Our approach is specific and based on three axes. First, our modeling approach is based on human movement science: we conduct challenging experiments focusing on the perception as well as on the motion involved in local interactions between walkers both using real and virtual set-ups. Second: we develop high-performance solutions for crowd simulation. Third, we develop solutions for realistic navigation in virtual world to enable interaction with crowds in Virtual Reality.

### 4.4. Motion Sensing of Human Activity

Recording human activity is a key point of many applications and fundamental works. Numerous sensors and systems have been proposed to measure positions, angles or accelerations of the user's body parts. Whatever the system is, one of the main problems is to be able to automatically recognize and analyze the user's performance according to poor and noisy signals. Human activity and motion are subject to variability: intravariability due to space and time variations of a given motion, but also inter-variability due to different styles and anthropometric dimensions. MimeTIC has addressed the above problems in two main directions.

Firstly, we have studied how to recognize and quantify motions performed by a user when using accurate systems such as Vicon (product of Oxford Metrics) or Optitrack (product of Natural Point) motion capture systems. These systems provide large vectors of accurate information. Due to the size of the state vector (all the degrees of freedom) the challenge is to find the compact information (named features) that enables the automatic system to recognize the performance of the user. Whatever the method used, finding these relevant features that are not sensitive to intra-individual and inter-individual variability is a challenge. Some researchers have proposed to manually edit these features (such as a Boolean value stating if the arm is moving forward or backward) so that the expertise of the designer is directly linked with the success ratio. Many proposals for generic features have been proposed, such as using Laban notation which was introduced to encode dancing motions. Other approaches tend to use machine learning to automatically extract these features. However most of the proposed approaches were used to seek a database for motions which properties correspond to the features of the user's performance (named motion retrieval approaches). This does not ensure the retrieval of the exact performance of the user but a set of motions with similar properties.

Secondly, we wish to find alternatives to the above approach which is based on analyzing accurate and complete knowledge on joint angles and positions. Hence new sensors, such as depth-cameras (Kinect, product of Microsoft) provide us with very noisy joint information but also with the surface of the user. Classical approaches would try to fit a skeleton into the surface in order to compute joint angles which, again, lead to large state vectors. An alternative would be to extract relevant information directly from the raw data, such as the surface provided by depth cameras. The key problem is that the nature of these data may be very different from classical representation of human performance. In MimeTIC, we try to address this problem in specific application domains that require picking specific information, such as gait asymmetry or regularity for clinical analysis of human walking.

# 4.5. VR and Sports

Sport is characterized by complex displacements and motions. These motions are dependent on visual information that the athlete can pick up in his environment, including the opponent's actions. Perception is thus fundamental to the performance. Indeed, a sportive action, as unique, complex and often limited in time, requires a selective gathering of information. This perception is often seen as a prerogative for action, it then takes the role of a passive collector of information. However, as mentioned by Gibson in 1979, the perception-action relationship should not be considered sequential but rather as a coupling: we perceive to act but we must act to perceive. There would thus be laws of coupling between the informational variables available in the environment and the motor responses of a subject. In other words, athletes have the ability to directly perceive the opportunities of action directly from the environment. Whichever school of thought considered, VR offers new perspectives to address these concepts by complementary using real time motion capture of the immersed athlete.

In addition to better understanding sports and interactions between athletes, VR can also be used as a training environment as it can provide complementary tools to coaches. It is indeed possible to add visual or auditory information to better train an athlete. The knowledge found in perceptual experiments can be for example used to highlight the body parts that are important to look at to correctly anticipate the opponent's action.

# 4.6. Interactive Digital Storytelling

Interactive digital storytelling, including novel forms of edutainment and serious games, provides access to social and human themes through stories which can take various forms and contains opportunities for

massively enhancing the possibilities of interactive entertainment, computer games and digital applications. It provides chances for redefining the experience of narrative through interactive simulations of computergenerated story worlds and opens many challenging questions at the overlap between computational narratives, autonomous behaviours, interactive control, content generation and authoring tools.

Of particular interest for the MimeTIC research team, virtual storytelling triggers challenging opportunities in providing effective models for enforcing autonomous behaviours for characters in complex 3D environments. Offering both low-level capacities to characters such as perceiving the environments, interacting with the environment and reacting to changes in the topology, on which to build higher-levels such as modelling abstract representations for efficient reasoning, planning paths and activities, modelling cognitive states and behaviours requires the provision of expressive, multi-level and efficient computational models. Furthermore virtual storytelling requires the seamless control of the balance between the autonomy of characters and the unfolding of the story through the narrative discourse. Virtual storytelling also raises challenging questions on the conveyance of a narrative through interactive or automated control of the cinematography (how to stage the characters, the lights and the cameras). For example, estimating visibility of key subjects, or performing motion planning for cameras and lights are central issues for which have not received satisfactory answers in the literature.

# 4.7. VR and Ergonomics

The design of workstations nowadays tends to include assessment steps in a Virtual Environment (VE) to evaluate ergonomic features. This approach is more cost-effective and convenient since working directly on the Digital Mock-Up (DMU) in a VE is preferable to constructing a real physical mock-up in a Real Environment (RE). This is substantiated by the fact that a Virtual Reality (VR) set-up can be easily modified, enabling quick adjustments of the workstation design. Indeed, the aim of integrating ergonomics evaluation tools in VEs is to facilitate the design process, enhance the design efficiency, and reduce the costs.

The development of such platforms asks for several improvements in the field of motion analysis and VR. First, interactions have to be as natural as possible to properly mimic the motions performed in real environments. Second, the fidelity of the simulator also needs to be correctly evaluated. Finally, motion analysis tools have to be able to provide in real-time biomechanics quantities usable by ergonomists to analyse and improve the working conditions.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

The results of the PhD thesis of Pierre Plantard leaded to the software "Kimea" with a national APP deposit. Faurecia company encouraged us to create a start-up company based on these results. Hence, we get two grants of the SATT "Ouest Valorisation" (total 300K) to transform the thesis prototype into a professionnal solution. The Kimea project has been granted several industrial and innovation prices (see below). A software engineer and an ergonomist have been recruited to create the original team of the future start-up (creation planned beginning of 2018). The software is based on several previously published works [14], and validated in actual industrial context [13]. Franck Multon will be scientific expert in the future start-up as a co-founder of the company.

#### 5.1.1. Awards

Kimea project has been granted by regional and national innovation committees:

- regional Pepite Tremplin competition by the Universities Bretagne Loire, October 2017,
- national Pepite Tremplin competition (53 projects granted among 700), November 2017,
- granted "Projet du futur" if the BPO foundation (10k), 18/10/2017,
- Top 500 national startup, "Hello Tomorrow" challenge, 25-27 october 2017.

# 6. New Software and Platforms

# 6.1. AsymGait

Asymmetry index for clinical gait analysis based on depth images

KEYWORDS: Motion analysis - Kinect - Clinical analysis

SCIENTIFIC DESCRIPTION: The system uses depth images delivered by the Microsoft Kinect to retrieve the gait cycles first. To this end it is based on a analyzing the knees trajectories instead of the feet to obtain more robust gait event detection. Based on these cycles, the system computes a mean gait cycle model to decrease the effect of noise of the system. Asymmetry is then computed at each frame of the gait cycle as the spatial difference between the left and right parts of the body. This information is computed for each frame of the cycle.

FUNCTIONAL DESCRIPTION: AsymGait is a software package that works with Microsoft Kinect data, especially depth images, in order to carry-out clinical gait analysis. First it identifies the main gait events using the depth information (footstrike, toe-off) to isolate gait cycles. Then it computes a continuous asymmetry index within the gait cycle. Asymmetry is viewed as a spatial difference between the two sides of the body.

- Participants: Edouard Auvinet and Franck Multon
- Contact: Franck Multon

## 6.2. Cinematic Viewpoint Generator

### KEYWORD: 3D animation

FUNCTIONAL DESCRIPTION: The software, developed as an API, provides a mean to automatically compute a collection of viewpoints over one or two specified geometric entities, in a given 3D scene, at a given time. These viewpoints satisfy classical cinematographic framing conventions and guidelines including different shot scales (from extreme long shot to extreme close-up), different shot angles (internal, external, parallel, apex), and different screen compositions (thirds,fifths, symmetric of di-symmetric). The viewpoints allow to cover the range of possible framings for the specified entities. The computation of such viewpoints relies on a database of framings that are dynamically adapted to the 3D scene by using a manifold parametric representation and guarantee the visibility of the specified entities. The set of viewpoints is also automatically annotated with cinematographic tags such as shot scales, angles, compositions, relative placement of entities, line of interest.

- Participants: Christophe Lino, Emmanuel Badier and Marc Christie
- Partners: Université d'Udine Université de Nantes
- Contact: Marc Christie

### 6.3. Directors Lens Motion Builder

KEYWORDS: Previzualisation - Virtual camera - 3D animation

FUNCTIONAL DESCRIPTION: Directors Lens Motion Builder is a software plugin for Autodesk's Motion Builder animation tool. This plugin features a novel workflow to rapidly prototype cinematographic sequences in a 3D scene, and is dedicated to the 3D animation and movie previsualization industries. The workflow integrates the automated computation of viewpoints (using the Cinematic Viewpoint Generator) to interactively explore different framings of the scene, proposes means to interactively control framings in the image space, and proposes a technique to automatically retarget a camera trajectory from one scene to another while enforcing visual properties. The tool also proposes to edit the cinematographic sequence and export the animation. The software can be linked to different virtual camera systems available on the market.

- Participants: Christophe Lino, Emmanuel Badier and Marc Christie
- Partner: Université de Rennes 1
- Contact: Marc Christie

# 6.4. Kimea

Kinect IMprovement for Egronomics Assessment

KEYWORDS: Biomechanics - Motion analysis - Kinect

SCIENTIFIC DESCRIPTION: Kimea consists in correcting skeleton data delivered by a Microsoft Kinect in an ergonomics purpose. Kimea is able to manage most of the occlultations that can occur in real working situation, on workstations. To this end, Kimea relies on a database of examples/poses organized as a graph, in order to replace unreliable body segments reconstruction by poses that have already been measured on real subject. The potential pose candidates are used in an optimization framework.

FUNCTIONAL DESCRIPTION: Kimea gets Kinect data as input data (skeleton data) and correct most of measurement errors to carry-out ergonomic assessment at workstation.

- Participants: Franck Multon, Hubert Shum and Pierre Plantard
- Partner: Faurecia
- Contact: Franck Multon
- Publications: Usability of corrected Kinect measurement for ergonomic evaluation in constrained environment - Validation of an ergonomic assessment method using Kinect data in real workplace conditions - Ergonomics Measurements using Kinect with a Pose Correction Framework - Filtered Pose Graph for Efficient Kinect Pose Reconstruction - Reliability of Kinect measurements for assessing the movement of operators in ergonomic studies

### 6.5. Populate

KEYWORDS: Behavior modeling - Agent - Scheduling SCIENTIFIC DESCRIPTION: The software provides the following functionalities:

- A high level XML dialect that is dedicated to the description of agents activities in terms of tasks and sub activities that can be combined with different kind of operators: sequential, without order, interlaced. This dialect also enables the description of time and location constraints associated to tasks.

- An XML dialect that enables the description of agent's personal characteristics.

- An informed graph describes the topology of the environment as well as the locations where tasks can be performed. A bridge between TopoPlan and Populate has also been designed. It provides an automatic analysis of an informed 3D environment that is used to generate an informed graph compatible with Populate.

- The generation of a valid task schedule based on the previously mentioned descriptions.

With a good configuration of agents characteristics (based on statistics), we demonstrated that tasks schedules produced by Populate are representative of human ones. In conjunction with TopoPlan, it has been used to populate a district of Paris as well as imaginary cities with several thousands of pedestrians navigating in real time.

FUNCTIONAL DESCRIPTION: Populate is a toolkit dedicated to task scheduling under time and space constraints in the field of behavioral animation. It is currently used to populate virtual cities with pedestrian performing different kind of activities implying travels between different locations. However the generic aspect of the algorithm and underlying representations enable its use in a wide range of applications that need to link activity, time and space. The main scheduling algorithm relies on the following inputs: an informed environment description, an activity an agent needs to perform and individual characteristics of this agent. The algorithm produces a valid task schedule compatible with time and spatial constraints imposed by the activity description and the environment. In this task schedule, time intervals relating to travel and task fulfillment are identified and locations where tasks should be performed are automatically selected.

- Participants: Carl-Johan Jorgensen and Fabrice Lamarche
- Contact: Fabrice Lamarche

# 6.6. The Theater

**KEYWORDS: 3D animation - Interactive Scenarios** 

FUNCTIONAL DESCRIPTION: The Theater is a software framework to develop interactive scenarios in virtual 3D environements. The framework provides means to author and orchestrate 3D character behaviors and simulate them in real-time. The tools provides a basis to build a range of 3D applications, from simple simulations with reactive behaviors, to complex storytelling applications including narrative mechanisms such as flashbacks.

- Participant: Marc Christie
- Contact: Marc Christie

## 6.7. CusToM

Customizable Toolbox for Musculoskeletal simulation

KEYWORDS: Biomechanics - Dynamic Analysis - Kinematics - Simulation - Mechanical multi-body systems SCIENTIFIC DESCRIPTION: The present toolbox aims at performing a motion analysis thanks to an inverse dynamics method.

Before performing motion analysis steps, a musculoskeletal model is generated. Its consists of, first, generating the desire anthropometric model thanks to models libraries. The generated model is then kinematical calibrated by using data of a motion capture. The inverse kinematics step, the inverse dynamics step and the muscle forces estimation step are then successively performed from motion capture and external forces data. Two folders and one script are available on the toolbox root. The Main script collects all the different functions of the motion analysis pipeline. The Functions folder contains all functions used in the toolbox. It is necessary to add this folder and all the subfolders to the Matlab path. The Problems folder is used to contain the different study. The user has to create one subfolder for each new study. Once a new musculoskeletal model is used, a new study is necessary. Different files will be automatically generated and saved in this folder. All files located on its root are related to the model and are valuable whatever the motion considered. A new folder will be added for each new motion capture. All files located on a folder are only related to this considered motion.

FUNCTIONAL DESCRIPTION: Inverse kinematics Inverse dynamics Muscle forces estimation External forces prediction

- Participants: Antoine Muller, Charles Pontonnier and Georges Dumont
- Contact: Antoine Muller

# 6.8. MotionGraphVR

KEYWORDS: Virtual reality - Motion capture - Movement analysis

FUNCTIONAL DESCRIPTION: MotionGraphVR is a tool enabling users to automatically create motion graphs in Unity. It is particularly targeting Virtual Reality applications, where with the development of Head Mounted Displays users are now unable to see their real body unless they use expensive motion capture system, or animation techniques (e.g., Inverse Kinematics) which suffer from a lack of visual realism. To lever these limitations, MotionGraphVR automatically builds a graph of human motions from a set of examples captured on a real actor, and identify which motion path is the graph is closest to the user's actions. Additionally, this plugin also provides analysing tools to allow developers of VR applications to visualise similarities between movements to use in their applications before seamlessly connecting them in Motion Graphs.

- Participants: Tiffany Luong, Ludovic Hoyet and Fernando Argelaguet Sanz
- Contact: Ludovic Hoyet

### 6.9. Platforms

#### 6.9.1. Immerstar Platform

Participants: Georges Dumont [contact], Ronan Gaugne, Anthony Sorel, Richard Kulpa.

With the two platforms of virtual reality, Immersia and Immermove, grouped under the name Immerstar, the team has access to high level scientific facilities. This equipment benefits the research teams of the center and has allowed them to extend their local, national and international collaborations. The Immerstar platform is granted by a Inria CPER funding for 2015-2019 that enables important evolutions of the equipment. In 2016, the first technical evolutions have been decided and, in 2017, these evolutions have been implemented. On one side, for Immermove, the addition of a third face to the immersive space, and the extension of the Vicon tracking system have been realized. And, on the second side, for Immersia, the installation of WQXGA laser projectors with augmented global resolution, of a new tracking system with higher frequency and of new computers for simulation and image generation.

# 7. New Results

### 7.1. Outline

In 2017, MimeTIC has maintained his activity in motion analysis, modelling and simulation. In motion analysis, we focused our efforts on three major points: 1) being able to simplify the calibration and simulation of customized musculoskeletal models of the subjects, 2) explore how visual perception act on collision avoidance in pedestrian locomotion with an extension to group behavior, and 3) adapt accurate analysis in real condition (industrial or clinical contexts) where measurement inaccuracies and easy-to-use constraints make it difficult to directly apply methods used in laboratories.

From a long time, MimeTIC has been promoting the idea of using Virtual Reality to train human performance. On the one hand, it leads to an efficient tradeoff between high control and naturalness of the situation. On the other hand, it raises several fundamental questions about the automatic evaluation of the performance of the user, and the transfer of the skills trained in VR to real practice. In 2017, we explored these two questions by 1) developping new automatic methods for users' performance recognition and evaluation, and 2) biofidelity of mass manipulation in VR using haptic interfaces.

In virtual cinematography, we applied the analysis/synthesis approach to extract and simulate film styles and narration. We also extended our previously defiend Toric Space for camera placement to drone toric space to control a group of drones filming the action of an actor to ensure the coverture of cinematographic distinct viewpoints.

## 7.2. Motion analysis

#### 7.2.1. Biomechanics for motion analysis-synthesis

Participants: Charles Pontonnier, Georges Dumont, Franck Multon, Antoine Muller, Diane Haering.

The PhD thesis of Antoine Muller defended on june, the 26 [1] aimed at democratizing the use of musculoskeletal analysis for a wide range of users. The work proposed contributions enabling better performances of such analyses and preserving accuracy, as well as contributions enabling an easy subject-specific model calibration. Firstly, in order to control the whole analysis process, the work is developed in a global approach of all the analysis steps: kinematics, dynamics [10] and muscle forces estimation. For all of these steps, quick analysis methods have been proposed. Particularly, a quick muscle force sharing problem resolution method [25] has been proposed, based on interpolated data. Moreover, a complete calibration process [24], based on classical motion analysis tools available in a biomechanical lab has been developed, based on motion capture and force platform data. Diane Haering, Inria Post-doctoral fellow at MimeTIC works on the the determination of maximal torque enveloppes of the elbow. These results could have a great potential of application to quantify the articular load during work tasks [19], to help calibrating muscle parameters into musculoskeletal simulations [5]. The method has been integrated in a more global subject specific calibration method [30]. This method could also be used to better represent musculoskeletal models as in [6].

Ana-Lucia Cruz-Ruiz was a PhD student from november 2013 to december 2016. The goal of this thesis was to define and evaluate muscle-based controllers for motion control. This PhD was related to the ANR Entracte project. She developed an original control approach to reduce the redudancy of the musculoskeletal system. A low-dimensional representation of control mechanisms in throwing motions from a variety of subjects and target distances was proposed. The control representation stands at the kinematic level in task and joint spaces respectively, and at the muscle activation level using the theory of muscle synergies. Representative features were chosen and extracted using factorization and clustering techniques from the muscle data leading to better represent mechanisms hidden behind such dynamical motions, and could offer a promising control representation for synthesizing motions with muscle-driven characters [3].

#### 7.2.2. Interactions between walkers

Participants: Anne-Hélène Olivier, Armel Crétual, Richard Kulpa, Sean Lynch, Laurentius Meerhoff.

Interaction between people, and especially local interaction between walkers, is a main research topic of MimeTIC. We propose experimental approaches using both real and virtual environments to study both perception and action aspects of the interaction. Our efforts to validate the virtual reality platform to study interactions was acknowledged by a publication in IEEE TVCG 2017 [11] and was presented in IEEE VR 2017 conference [26]. Using the VR platform, we investigated the nature of visual information that is used for a collision free interaction. We aimed to manipulate the nature of visual information in two forms, global and local information appearances. The obstacle was presented with one of five virtual appearances, associated to global motion cues (i.e., a cylinder or a sphere), or local motion cues (i.e., only the legs or the trunk). A full body virtual walker, showing both local and global motion cues, used as a reference condition. The final crossing distance was affected by the global motion appearances, however, appearance had no qualitative effect on motion adaptations. These findings contribute towards further understanding what information people use when interacting with others. This work was published in TVCG 2017 [7] and presented as a poster in the ACAPS 2017 Conference[36]. This year, we also developed new experiments in our immersive platform. We designed a study to investigate the effect of gaze interception during collision avoidance between two walkers. In such a situation, mutual gaze can be considered as a form of nonverbal communication. Additionally, gaze is believed to detail future path intentions and to be part of the nonverbal negotiation to achieve avoidance collaboratively. We considered an avoidance task between a real subject and a virtual human character and studied the influence of the character's gaze direction on the avoidance behaviour of the participant. Virtual reality provided us with an accurate control of the situation: seventeen participants were immersed in a virtual environment, instructed to navigate across a virtual space using a joystick and to avoid a virtual character that would appear from either side. The character would either gaze or not towards the participant. Further, the character would either perform or not a reciprocal adaptation of its trajectory to avoid a potential collision with the participant. The findings of this paper were that during an orthogonal collision avoidance task, gaze behaviour did not influence the collision avoidance behaviour of the participants. Further, the addition of reciprocal collision avoidance with gaze did not modify the collision behaviour of participants. These results suggest that for the duration of interaction in such a task, body motion cues were sufficient for coordination and regulation. We discuss the possible exploitation of these results to improve the design of virtual characters for populated virtual environments and to interact with users. These results were presented to the AFRV 2017 conference [33] and submitted to IEEE VR 2018 conference.

We also provide lot of efforts to investigate, in collaboration with Julien Pettré from Inria Lagadic team, the process involved in the selection of interactions within our neighbourood. Considering the complex case of multiple interactions, we performed experiments in real conditions where a participant walked across a room whilst either one (i.e., pairwise) or two (i.e., group) participants crossed the room perpendicularly. By comparing these pairwise and group interactions, we assessed whether a participant avoids two upcoming collisions simultaneously, or as sequential pairwise interactions. Furthermore, in the group trials we varied the relative position of the two participants that crossed the trajectory of the other. This allowed us to change the affordance of passing through or around (i.e., its 'pass-ability'). Results showed that in the group trials, participants consistently avoided collision with lower risks of impending collision (as quantified by the future distance of closest approach) in the group compared to the pairwise trials. This implies that a participant – to some extent – interacted simultaneously with two other participants. Furthermore, we analysed in the group trials how the 'pass-ability' evolved over time. Results indicated that the affordance of passing through or around was already established early in the interaction. This shows that participants are susceptible to the affordance of passing through a gap between others. We concluded that pedestrians are able to interact with two other walkers simultaneously, rather than treating each interaction in sequence. These results were presented at the ICPA 2017 conference [21].

Finally, we continue working on the interaction between a walker and a moving robot. This work was performed in collaboration with Philippe Souères and Christian Vassallo (LAAS, Toulouse). The development of Robotics accelerated these recent years, it is clear that robots and humans will share the same environment in a near future. In this context, understanding local interactions between humans and robots during locomotion tasks is important to steer robots among humans in a safe manner. Our work is a first step in this direction. Our goal is to describe how, during locomotion, humans avoid collision with a moving robot. We just published in Gait and Posture our results on collision avoidance between participants and a non-reactive robot (we wanted to avoid the effect of a complex loop by a robot reacting to participants' motion). Our objective was to determine whether the main characteristics of such interaction preserve the ones previously observed: accurate estimation of collision risk, anticipated and efficient adaptations. We observed that collision avoidance between a human and a robot has similarities with human-human interactions (estimation of collision risk, anticipation) but also leads to major differences [17]. Humans preferentially give way to the robot, even if this choice is not optimal with regard to motion adaptation to avoid the collision. In this new study, we considered the situation where the robot was reactive to the walker's motion. First of all, it results that humans have a good understanding of the robot behavior and their reaction are smoother and faster with respect to the case with a non-collaborative robot. Second, humans adapt similarly to human-human study and the crossing order is respected in almost all cases. These results have strong similarities with the ones observed with two humans crossing each other.

#### 7.2.3. New automatic methods to assess motion in industrial contexts based on Kinect

Participants: Franck Multon, Georges Dumont, Charles Pontonnier, Pierre Plantard, Antoine Muller.

Recording human activity is a key point of many applications and fundamental works. Numerous sensors and systems have been proposed to measure positions, angles or accelerations of the user's body parts. Whatever the system is, one of the main challenge is to be able to automatically recognize and analyze the user's performance according to poor and noisy signals. Hence, recognizing and measuring human performance are important scientific challenges especially when using low-cost and noisy motion capture systems. MimeTIC has addressed the above problems in two main application domains. In this section, we detail the ergonomics application of such an approach. Firstly, in ergonomics, we explored the use of low-cost motion capture systems (i.e., a Microsoft Kinect) to measure the 3D pose of a subject in natural environments, such as on a workstation, with many occlusions and inappropriate sensor placements. Predicting the potential accuracy of the measurement for such complex 3D poses and sensor placements is challenging with classical experimental setups. After having evaluated the actual accuracy of the pose reconstruction method delivered by the Kinect, we have identified that occlusions were a very important problem to solve in order to obtain reliable ergonomic assessments in real cluttered environments. To this end, we developed an approach to deal with long occlusions that occur in real manufacturing conditions. This approach is based on a structured database of examples (named filtered pose graph) that enables real-time correction of Kinect skeleton data [14].

This method has been applied to a complete ergonomic process outputting RULA scores based on the reconstructed and corrected poses. We challenged this method with a reference motion capture system in laboratory conditions [15]. To this end we compared joint angles and RULA scores obtained with our system and a reference Vicon mocap system in various conditions (with and without occlusions). The results show a very good accordance between manually tuned RULA scores given by experts and those computed by

the automatic system. These results demonstrate that it could be used in industrial context to support the ergonomists decision-making process.

This year we also extended this work to evaluate if corrected data enabled us to estimate reliable joint torques using inverse dynamics to provide new information to ergonomic assessment [12]. Indeed, joint torques and forces are relevant quantities to estimate the biomechanical constraints of working tasks in ergonomics. However, inverse dynamics requires accurate motion capture data, which are generally not available in real manufacturing plants. Markerless and calibrationless measurement systems based on depth cameras, such as the Microsoft Kinect, are promising means to measure 3D poses in real time, such as using our corrected Kinect approach. Thus, we evaluated the reliability of an inverse dynamics method based on this corrected skeleton data and its potential use to estimate joint torques and forces in such cluttered environments. To this end, we compared the calculated joint torques with those obtained with a reference inverse dynamics method based on an optoelectronic motion capture system. Results show that the Kinect skeleton data enabled the inverse dynamics process to deliver reliable joint torques in occlusion-free (r=0.99 for the left shoulder elevation) and occluded (r=0.91 for the left shoulder elevation) environments. However, differences remain between joint torques estimations. Such reliable joint torques open appealing perspectives for the use of new fatigue or solicitation indexes based on internal efforts measured on site. The study demonstrates that corrected Kinect data could be used to estimate internal joint torques, using an adapted inverse dynamics method. The method could be applied on-site because it can handle some cases with occlusions. The resulting Kinect-based method is easy-to-use, real-time and could assist ergonomists in risk evaluation on site.

This work was partially funded by the Faurecia company through a Cifre convention.

## 7.2.4. Clinical gait assessment based on Kinect data

#### Participant: Franck Multon.

In clinical gait analysis, we proposed a method to overcome the main limitations imposed by the low accuracy of the Kinect measurements in real medical exams. Indeed, inaccuracies in the 3D depth images lead to badly reconstructed poses and inaccurate gait event detection. In the latter case, confusion between the foot and the ground leads to inaccuracies in the foot-strike and toe-off event detection, which are essential information to get in a clinical exam. To tackle this problem we assumed that heel strike events could be indirectly estimated by searching for the extreme values of the distance between the knee joints along the walking longitudinal axis. As Kinect sensor may not accurately locate the knee joint, we used anthropometrical data to select a body point located at a constant height where the knee should be in the reference posture. Compared to previous works using a Kinect, heel strike events and gait cycles are more accurately estimated, which could improve global clinical gait analysis frameworks with such a sensor. Once these events are correctly detected, it is possible to define indexes that enable the clinician to have a rapid state of the quality of the gait. We therefore proposed a new method to assess gait asymmetry based on depth images, to decrease the impact of errors in the Kinect joint tracking system. It is based on the longitudinal spatial difference between lower-limb movements during the gait cycle. The movement of artificially impaired gaits was recorded using both a Kinect placed in front of the subject and a motion capture system. The proposed longitudinal index distinguished asymmetrical gait, while other symmetry indices based on spatiotemporal gait parameters failed using such Kinect skeleton measurements. This gait asymmetry index measured with a Kinect is low cost, easy to use and is a promising development for clinical gait analysis.

This method has been challenged with other classical approaches to assess gait asymmetry using either cheap Kinect data or Vicon data. We demonstrate the superiority of the approach when using Kinect data for which traditional approaches failed to accurately detect gait asymmetry. It has been validated on healthy subjects who were forced to walk with a 5cm sole placed below each foot alternatively. In 2017 [2], we compared the results obtained with the famous Constant Relative Phase (CRP) that aims at quantifying within-stride asymmetry index. CRP requires noise-free and accurate motion capture, which is difficult to obtain in clinical settings. As our index, the Longitudinal Asymmetry Index (ILong), is obtained using data from a low-cost depth camera (Kinect) (depth images averaged over several gait cycles), rather than derived joint positions or angles, we checked that it could deliver more reliable asymmetry information within gait, compared to CRP. Hence, this

study aimed to evaluate (1) the validity of CRP computed with Kinect, (2) the validity and sensitivity of ILong for measuring gait asymmetry based solely on data provided by a depth camera, (3) the clinical applicability of a posteriorly mounted camera system to avoid occlusion caused by the standard front-fitted treadmill consoles and (4) the number of strides needed to reliably calculate ILong. the results show that CRP based on times derivatives of joint angles failed to detect gait asymmetry, when using Kinect data. However, our index, ILong, detected this disturbed gait reliably and could be computed from a posteriorly placed Kinect without loss of validity. A minimum of five strides was needed to achieve a correlation coefficient of 0.9 between standard MBS and low-cost depth camera based ILong. ILong provides a clinically pragmatic method for measuring gait asymmetry, with application for improved patient care through enhanced disease, screening, diagnosis and monitoring.

This work has been done in collaboration with the MsKLab from Imperial College London, to design new gait asymmetry indexes that could be used in daily clinical analysis.

#### 7.2.5. Biomechanical analysis of tennis serve

Participants: Caroline Martin, Richard Kulpa, Benoit Bideau, Pierre Touzard.

Following the previous studies we made on tennis serve, we were able to evaluate the link between performance and risk of injuries. To go further, we made new experiments on top-level young French players (between 12 up to 18 years old) to quantify the motor technical errors made (kinematics) and the impact on the risk of injury (dynamics). This experiments are part of a collaboration with the FFT (French Tennis Federation). We recently validated that the Waiter's serve implies higher risk of injuries [28]. It is a movement that was know by the coaches as not productive and risky but it was never validated.

# 7.3. Virtual human simulation

### 7.3.1. Novel Distance Geometry based approaches for Human Motion Retargeting

Since September 2016, Antonio Mucherino has a half-time Inria detachment in the MimeTIC team, in order to collaborate on exploring distance geometry-based problems in representing and editing human motion.

In this context, an extension of a distance geometry approach to dynamical problems was proposed in [23], and we co-supervised Antonin Bernardin for his Master thesis, which focused on applying such extended approach for retargeting human motions. In character animation, it is often the case that motions created or captured on a specific morphology need to be reused on characters having a different morphology. However, specific relationships such as body contacts or spatial relationships between body parts are often lost during this process, and existing approaches typically try to determine automatically which body part relationships should be preserved in such animation. Instead, we proposed a novel frame-based approach to motion retargeting [18], [22] which relies on a normalized representation of all the body joints distances to encompass all the relationships existing in a given motion. In particular, we proposed to abstract postures by computing all the inter-joint distances of each animation frame and to represent them by Euclidean Distance Matrices (EDMs). Such EDMs present the benefits of capturing all the subtle relationships between body parts, while being adaptable through a normalization process to create a morphology independent distance-based representation. Finally, they can also be used to efficiently compute retargeted joint positions best satisfying newly imposed distances. We demonstrated that normalized EDMs can be efficiently applied to a different skeletal morphology by using a dynamical distance geometry approach, and presented results on a selection of motions and skeletal morphologies.

In parallel, in collaboration with national (LIX, École Polytechnique, Palaiseau) and international partners, we have been working for improving the performances of existing algorithms for distance geometry, independently from the considered application. In [4], we analyzed the main causes for the approach to fail to provide accurate solutions in cases where interval distances are provided (instead of unique distance values), and we proposed some possible strategies to detect such situations. In [27], we presented a linear optimization problem for a common pre-processing step in distance geometry: the one of identifying a special vertex order allowing to discretize the solution search space.

# 7.4. Human motion in VR

#### 7.4.1. Motion recognition and classification

Participants: Franck Multon, Richard Kulpa, Yacine Boulahia.

Action recognition based on human skeleton structure represents nowadays a prospering research field. This is mainly due to the recent advances in terms of capture technologies and skeleton extraction algorithms. In this context, we observed that 3D skeleton-based actions share several properties with handwritten symbols since they both result from a human performance. We accordingly hypothesize that the action recognition problem can take advantage of trial and error approaches already carried out on handwritten patterns. Therefore, inspired by one of the most efficient and compact handwriting feature-set, we proposed a skeleton descriptor referred to as Handwriting-Inspired Features. First of all, joint trajectories are preprocessed in order to handle the variability among actor's morphologies. Then we extract the HIF3D features from the processed joint locations according to a time partitioning scheme so as to additionally encode the temporal information over the sequence. Finally, we used Support Vector Machine (SVM) for classification. Evaluations conducted on two challenging datasets, namely HDM05 and UTKinect, testify the soundness of our approach as the obtained results outperform the state-of-the-art algorithms that rely on skeleton data [32].

This work has been carried-out in collaboration with the IRISA Intuidoc team, with Yacine Boulahia who is a co-supervised PhD student with Eric Anquetil.

### 7.4.2. Automatic evaluation of sports gesture

Participants: Richard Kulpa, Marion Morel.

Automatically evaluating and quantifying the performance of a player is a complex task since the important motion features to analyze depend on the type of performed action. But above all, this complexity is due to the variability of morphologies and styles of both the experts who perform the reference motions and the novices. Only based on a database of experts' motions and no additional knowledge, we propose an innovative 2-level DTW (Dynamic Time Warping) approach to temporally and spatially align the motions and extract the imperfections of the novice's performance for each joints [9]. We applied our method on tennis serve and karate katas [8].

### 7.4.3. Biofidelity in VR

Participants: Hilt Simon, Charles Pontonnier, Georges Dumont.

Recording human activity is a key point of many applications and fundamental works. Numerous sensors and systems have been proposed to measure positions, angles or accelerations of the user's body parts. Whatever the system is, one of the main challenge is to be able to automatically recognize and analyze the user's performance according to poor and noisy signals. Hence, recognizing and measuring human performance are important scientific challenges especially when using low-cost and noisy motion capture systems. MimeTIC has addressed the above problems in two main application domains. In this section, we detail the ergonomics application of such an approach. Firstly, in ergonomics, we explored the use of low-cost motion capture systems (i.e., a Microsoft Kinecte of geometrical and mechanical characteristics of the haptic device. Uncertainties on friction coefficients within the model are tuned thanks to an experimental protocol enabling a subjective comparison between real and virtual manipulations of a low mass object. The compensation of friction on the first and second axes of the haptic interface showed significant improvement of both realism and perceived load [20].

# 7.5. Digital story telling

### 7.5.1. Analysis of Film Style

Participants: Marc Christie, Hui-Yin Wu, Christophe Lino, Quentin Galvane.

We have designed and made available an open database of annotated film clips together with an analysis of elements of film style related to how the shots are composed, how the transitions are performed between shots

and how the shots are sequenced to compose a film unit [29]. The purpose is to initiate a shared repository pertaining to elements of film style which can be used by computer scientists and film analysts alike. Though both research communities rely strongly on the availability of such information to foster their findings, current databases are either limited to low-level features (such as shots lengths, color and luminance information), contain noisy data, or are not available to the communities. The data and analysis we provide open exciting perspectives as to how computational approaches can rely more thoroughly on information and knowledge extracted from existing movies, and also provide a better understanding of how elements of style are arranged to construct a consistent message.

#### 7.5.2. Film Editing Patterns: Thinking like a Director

Participants: Marc Christie, Hui-Yin Wu.

We have introduced *Film Editing Patterns (FEP)*, a language to formalize film editing practices and stylistic choices found in movies. FEP constructs are constraints expressed over one or more shots from a movie sequence [34] that characterize changes in cinematographic visual properties such as shot size, region, angle of on-screen actors.

We have designed the elements of the FEP language, then introduced its usage in annotated film data, and described how it can support users in the creative design of film sequences in 3D. More specifically: (i) we proposed the design of a tool to craft edited filmic sequences from 3D animated scenes that uses FEPs to support the user in selecting camera framings and editing choices that follow certain best practices used in cinema; (ii) we conducted an evaluation of the application with professional and non-professional filmmakers. The evaluation suggested that users generally appreciate the idea of FEP, and that it can effectively help novice and medium experienced users in crafting film sequences with little training and satisfying results.

#### 7.5.3. Directing Cinematographic Drones

#### Participant: Marc Christie.

We have designed a set of high-level tools for filming dynamic targets with quadrotor drones. To this end, we proposed a specific camera parameter space (the Drone Toric space) together with interactive on-screen viewpoint manipulators compatible with the physical constraints of a drone. We then designed a real-time path planning approach in dynamic environments which ensures both cinematographic properties in viewpoints along the path and ensures the feasibility of the path by a quadrotor drone. We finally have demonstrated how the Drone Toric Space can be combined with our path planning technique to coordinate positions and motions of multiple drones around dynamic targets to ensure the coverture of cinematographic distinct viewpoints. The proposed research prototypes have been evaluation by an experienced drone pilot and filmmaker, as well as by non-experts users. Not only does the tool demonstrates it's usability for everyday recording of aethetic camera motions.

# 8. Bilateral Contracts and Grants with Industry

# 8.1. Bilateral Contracts with Industry

### 8.1.1. BPI-PCR Robo-KII

Participant: Armel Crétual [contact].

This contract has started in February 2017 and will last in October 2018. In M2S, it involves two permanent members of MimeTic team, Armel Crétual and Franck Multon, and two engineers, Antoine Marin (18 months grant) and Brice Bouvier (10 months grant).

This project is a collaboration between BA Healthcare and M2S lab. It aims at developing a robotics platform to allow physicians to start gait rehabilitation as soon as possible, even before patients are able to maintain upright posture alone. The usual way to perform such rehab sessions is to make the patient walk on a treadmill benefiting from a harness to prevent patient from falling. The two main limits of this approach are that:

- only straightforward at constant speed gaits are feasible whereas falling risks are much higher when modifying speed or turning
- walking on a treadmill when motor abilities are very affected can be challenging and can generate strong apprehension.

In a previous project, Robo-K, ended in september 2016, BA Healthcare has developed a first prototype of a mobile robot which strongly modified the approach: the harness is mobile and follows the patient displacement. In this way, the patient walks on the ground at his/her desired speed and the physician can include curved trajectories in the rehab process.

The main novelty of Robo-KII project is to implement a biofeeedback system onto the robotics platform to reinforce rehab sessions. Closely working with physicians from two PMR services, CHU Rennes and Kerpape center, we intend to define the optimal feedback to be given to the patients and to measure the corresponding gait parameters thanks to depth cameras mounted on the robot.

# 8.2. Bilateral Grants with Industry

Participants: Marc Christie, Christophe Lino.

Bilateral contract with Technicolor on empowering drones with cinematographics knowledge. Participants: Philippe Guillotel, Julien Fleureau, Quentin Galvane. Amount 25kE. Duration 24 months.

# 9. Partnerships and Cooperations

# 9.1. Regional Initiatives

- SATT "Ouest valorisation" grant for the maturation of the Kimea software and projet (Franck Multon and Pierre Plantard). 12 months of three full-time people 300K€. Creation of the start-up company planned beginning of 2018.
- SATT "Ouest valorisation" grant for the maturation of the Populate software (Fabrice Lamarche). One full-time engineer (2017-2018).

# 9.2. National Initiatives

### 9.2.1. ANR

### 9.2.1.1. Cineviz

Participants: Marc Christie [contact], Christophe Lino, Quentin Galvane, Hui-Yin Wu.

Cineviz is a 3-year ANR LabCom project (2016-2019). Amount: 300kE. Parnters: SolidAnim, UR1.

The project is a bilateral collaboration with the SolidAnim company. The objective is to jointly progress on the design and implementation of novel tools for the preproduction in the film industry. The project will address the challenges related to (i) proposing expressive framing tools, (ii) integrating the technical aspects of shooting (how to place the cameras, lights, green sets) directly at the design stage), and (iii) novel interaction metaphors for designing and controlling the staging of lights in preproduction, using an example-based approach.

#### 9.2.1.2. Entracte

**Participants:** Charles Pontonnier [contact], Georges Dumont, Franck Multon, Pierre Plantard, Ana Lucia Cruz Ruiz, Antoine Muller, Anthony Sorel, Nicolas Bideau, Richard Kulpa.

The ANR project ENTRACTE is a collaboration between the Gepetto team in LAAS, Toulouse (head of the project) and the Inria/MimeTIC team. The project started in November 2013 and ended in August 2017. The purpose of the ENTRACTE project is to address the action planning problem, crucial for robots as well as for virtual human avatars, in analyzing human motion at a biomechanical level [16] and in defining from this analysis bio-inspired motor control laws and bio-inspired paradigms for action planning. The project is launched since november 2013 and Ana Lucia Cruz Ruiz, who has been recruited as a PhD student since this date, defended her thesis on muscle-based control based on synergies last year.

#### 9.2.2. National scientific collaborations

#### 9.2.2.1. Cavaletic

Participant: Franck Multon [contact].

The Cavaletic collaborative project is leaded by University Bretagne Sud and also involves University Rennes2 (CREAD Lab.). It has been funded by the National IFCE (Institut Français du Cheval et de l'Equitation) in order to develop and evaluate technological assistance in horse riding learning, thanks to a user-centered approach. MimeTIC is involved in measuring expert and non-expert horse riders' motions in standardized situations in order to develop metrics to measure riders' performance. It will be used to develop a technological system embedded on users to evaluate their performance and provide them with real-time feedback to correct potential errors.

### 9.2.2.2. FFT

Participants: Richard Kulpa [contact], Benoit Bideau, Pierre Touzard.

An exclusive contract has been signed between the M2S laboratory and the French Federation of Tennis for three years. The goal is to perform biomechanical analyses of 3D tennis serves on a population of 40 players of the Pôle France. The objective is to determine the link between injuries and biomechanical constraints on joints and muscles depending on the age and gender of the players. At the end, the goal is to evaluate their load training.

#### 9.2.2.3. gDGA

Participants: Antonio Mucherino [contact], Ludovic Hoyet, Franck Multon.

gDGA (generalization of the Distance Geometry and its Applications) is a INS2I/CNRS PEPS project involving local and national partners. Distance geometry can nowadays be seen as a classical problem in operational research, having a wide range of applications. The main aim of this interdisciplinary project is to extend the definition and the range of applicability of distance geometry. In particular, our main interest is on dynamical problems, motivated by a certain number of applications of interest, including interaction motion adaptation, the simulation of crowd behaviours, and the conception of modern recommender systems. The classical application of distance geometry arising in the biological field is also taken into consideration. The necessity of a strong computational power for the considered applications motivates the need of implementing our algorithms in environments capable of exploiting the resources on GPU cards.

#### 9.2.2.4. IRMA

Participants: Ronan Gaugne [contact], Georges Dumont.

The IRMA project is an Imag'In project funded by CNRS which aims at developping innovative methodologies for research in the field of cultural heritage based on the combination of medical imaging technologies and interactive 3D technologies (virtual reality, augmented reality, haptics, additive manufacturing). It relies on close collaborations with the National Institute of Preventive Archaeological Research (Inrap), the Research Center Archaeology, and History Archéosciences (CReAAH UMR 6566) and the company Image ET. The developed tools are intended for cultural heritage professionals such as museums, curators, restorers, and archaeologists. We focus on a large number of archeological artefacts of different nature, and various time periods (Paleolithic, Mesolithic, and Iron Age Medieval) from all over France. We can notably mention the oldest human bones found in Brittany (clavicle Beg Er Vil), a funeral urn from Trebeurden (22), or a Bronze Cauldron from a burial of the Merovingian necropolis "Crassés Saint-Dizier" (51). This project involves a strong collaboration with members of the team Hybrid (Valérie Gouranton, Bruno Arnaldi and Jean-Baptiste Barreau), Théophane Nicolas (Inrap/UMR Trajectoires), Quentin Petit (SED Inria Rennes), and Grégor Marchand (CNRS/UMR CREAAH).

### 9.2.3. ADT: Immerstar

Participants: Franck Multon, Georges Dumont [contact], Ronan Gaugne.

The ADT-Immerstar is driven by the SED and aims at developing new tools and facilities for the scientific community in order to develop demos and use the two immersive rooms in Rennes: Immersia and Immermove. The engineer (Quentin Petit, SED) has the responsibility of homogenizing the software modules and development facilities in each platform, of installing new upgrades and of developping collaborative applications between the two sites.

### 9.2.4. PRE

Participants: Franck Multon [contact], Ludovic Hoyet, Antonio Mucherino.

The Inria PRE projet entitled "Smart sensors and novel motion representation breakthrough for human performance analysis" aims at designing a new description for human motion in order to automatically capture, measure and transfer the intrinsic constraints of human motion. Current approached consisted in manually editing the constraints associated with a motion, to use classical skeleton representation with joint angles based on direct or indirect measurements, and then perform inverse kinematics to fulfill these constraints. We aim at designing a new representation to simplify this process pipeline and make it automatic, together with relevant motion sensors that could provide enough information to automatically extract these intrinsic constraints. To this end, this project has been jointly proposed with the Inria CAIRN team, which develops sensors based on joint orientations and distances between sensors. We aim at extending this type of device to measure new types of information that would help to simplify the above mentionned pipeline. A postdoc arrived in November 2016 to jointly work with CAIRN. We also involved Hubert Shum from Northumbria University to link this project with the long-term collaboration with Dr. Shum about this type of problem. ype of problems.

# 9.3. International Initiatives

### 9.3.1. FORMOSA

Title: Fostering Research on Models for Storytelling Applications

International Partner (Institution - Laboratory - Researcher):

NCCU (Taiwan) - Intelligent Media Lab (IML) - Tsai-Yen Li

Start year: 2016

See also: http://www.irisa.fr/mimetic/GENS/mchristi/EA-FORMOSA/

Interactive Storytelling is a new media which allows users to alter the content and outcome of narratives through role-playing and specific actions. With the quality, the availability and reasonable costs of display technologies and 3D interaction devices on one side, and the accessibility of 3D content creation tools on the other, this media is taking a significant share in entertainment (as demonstrated by the success of cinematographic games such as Heavy Rain or Beyond: two souls). These advances push us to re-think the way narratives are traditionally structured, explore new interactive modalities and provide new interactive cinematographic experiences. As a sequel of the first associate team FORMOSA 1, we propose to address new challenges pertained to interactive storytelling such as the use of temporal structures in narratives, interaction modalities and their impact in terms of immersion, and the adaptation of cinematographic real data to 3D environments. To achieve these objectives, the associate team will rely on the complementary skills of its partners and on the co-supervision of students.

### 9.3.2. SIMS

Title: REal data against crowd SImulation AlgorithMS

International Partner (Institution - Laboratory - Researcher):

University of North Carolina at Chapel Hill (United States) - GAMMA Research Group (GAMMA) - Ming LIN

Start year: 2015

See also: http://www.irisa.fr/mimetic/GENS/jpettre/EASIMS/easims.html

RE-SIMS aims at gathering the best international research teams working on crowd simulation to allow significant progresses on the level of realism achieved by crowd simulators. To this end, RE-SIMS aims at improving methods for capturing crowd motion data that describe real crowd behaviors, as well as by improving data assimilation techniques.

In this renewal, RE-SIMS extends the previous SIMS partnership and follows a multidiciplinary direction.

### 9.3.3. Inria International Partners

9.3.3.1. Informal International Partners

- Dr. Edouard Auvinet, Imperial College London, UK (collaboration with Franck Multon)
- Dr. Huber Shum, Northumbria University, Newcastle, UK (collaboration with Franck Multon and Ludovic Hoyet, with joint papers and supervision)
- Dr. Rachel McDonnell, Trinity College Dublin, Ireland (on-going collaboration with Ludovic Hoyet, including a 6-month internship from one of her PhD student in Rennes)
- Prof. Carol O'Sullivan, Trinity College Dublin, Ireland (on-going collaboration with Ludovic Hoyet)
- Prof. Carlile Lavor, UNICAMP, Campinas, Sao Paulo, Brazil (collaboration with Antonio Mucherino)
- Dr. Douglas S. Gonçalves, Federal University of Santa Catarina, Florianpolis, Brazil (collaboration with Antonio Mucherino)
- Jung-Hsin Lin, Academia Sinica, Taipei, Taiwan (collaboration with Antonio Mucherino)

# 9.4. International Research Visitors

#### 9.4.1. Visits of International Scientists

#### 9.4.1.1. Professors and associate professors

- Victoria Interrante, Professor, Department of Computer Science and Engineering, University of Minnesota USA, December 8th, 2017
- Michael Cinelli, Associate Professor, Kinesiology and Physical Education, Wilfrid Laurier University, Canada, June 2017

9.4.1.2. Internships

• Emma Carrigan, Trinity College Dublin, Ireland (PhD supervisor: Dr. Rachel McDonnell), 6-month internship in collaboration with Technicolor (Quentin Avril), Jan. to June 2017.

# **10. Dissemination**

# **10.1. Promoting Scientific Activities**

### 10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- Anne-Hélène Olivier, Workshop VHCIE 2017, IEEE VR 2017, Los Angles, United-States, March 2017
- Anne-Hélène Olivier, Workshop Interactions, Rennes, June 2017

10.1.1.2. Member of the Organizing Committees

Ludovic Hoyet, co-organiser of the French Computer Graphics - Virtual Reality days, October 2017, Rennes, France

Marc Christie, steering committee of Motion in Games 2017

#### 10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

Ludovic Hoyet: co-chair of French Computer Graphics days 2017, October 2017, Rennes, France Antonio Mucherino: Workshop on Computational Optimization (WCO17), in the framework of the "Federated Conference on Computer Science and Information Systems" (FedCSIS17), co-chair with Stefka Fidanova and Daniela Zaharie Web: https://fedcsis.org/2017/wco

10.1.2.2. Member of the Conference Program Committees

Ludovic Hoyet, ACM Motion in Games MIG 2017, Barcelona, Spain, Nov. 2017

Ludovic Hoyet, ACM Symposium on Applied Perception 2017, Cottbus, Germany, Sept. 2017

Ludovic Hoyet, International Conference on Computer Graphics Theory and Applications 2018, Maderia, Portugal, Jan. 2018

Anne-Hélène Olivier, ACM Motion in Games MIG 2017, Barcelona, Spain, October 2017

Anne-Hélène Olivier, IEEE VR 2018 TVCG Conference paper, Reutlingen, Germany, March 2018

Franck Multon, ACM Motion in Games MIG 2017, Barcelona, Spain, Nov. 2017

Franck Multon, Computer Animation and Social Agents CASA 2017, Seoul, Korea, May 2017

Franck Multon, IEEE Conference on Automatic Face and Gesture Recognition FG'2017, Washington DC, US, June 2017

Franck Multon, Affective Computing and Intelligent Interaction ACII 2017, San Antonio, United States, October 2017

Richard Kulpa, International Conference on Computer Graphics Theory and Applications (GRAPP), 2017

Marc Christie, Motion In Games (MIG) 2017

### 10.1.2.3. Reviewer

Ludovic Hoyet, ACM Siggraph, Los Angeles, United-States, July 2017

Ludovic Hoyet, IEEE VR 2018, Reutlingen, Germany, March 2018

Ludovic Hoyet, Pacific Graphics, Taipei, Taiwan, October 2017

Anne-Hélène Olivier, IEEE VR 2018 TVCG Conference paper, Reutlingen, Germany, March 2018

Anne-Hélène Olivier, ACM Motion in Games MIG 2017, Barcelona, Spain, October 2017

Anne-Hélène Olivier, ACM VRST 2017, Gothenburg, Sweden, November 2017

Richard Kulpa, IEEE International Conference on Automatic Face and Gesture Recognition, 2018 Richard Kulpa, Siggraph CHI, 2018

Richard Kulpa, GRAPP, 2018

Marc Christie Eurographics 2017

Marc Christie ACM Siggraph CHI 2017

Marc Christie Siggraph 2017

Marc Christie IEEE VR 2017

Franck Multon IEEE VR 2017 Franck Multon ACM VRST 2017 Franck Multon ACM MIG 2017 Franck Multon ACM VRST 2017

### 10.1.3. Journal

#### 10.1.3.1. Member of the Editorial Boards

Franck Multon, Presence, MIT Press

Franck Multon, Computer Animation and Virtual Worlds CAVW, John Wiley

Antonio Mucherino, Guest Editor of Optimization Letters, Springer

Armel Crétual, Editorial board of Journal of Electromyography and Kinesiology

Marc Christie, Associate Editor of the Visual Computer

#### 10.1.3.2. Reviewer - Reviewing Activities

Franck Multon, Medical & Biological Engineering & Computing, Computer and Graphics, IEEE Journal of Biomedical and Health Informatics, IEEE Transactions on Visualization and Computer Graphics, Journal of Computer Science and Technology, Computer Animation and Virtual Worlds, ROBOMECH journal

Ludovic Hoyet, ACM Transactions on Graphics, IEEE Transactions on Visualization and Computer Graphics, Computers & Graphics, Computer Graphics Forum

Anne-Hélène Olivier, Gait and Posture, Motor Control

Antonio Mucherino, Journal of Global Optimization, Springer, Discrete Applied Mathematics, Elsevier

Armel Crétual, Journal of Electromyography and Kinesiology, Clinical Biomechanics, Gait & Posture, Journal of Orthopaedic Research, Medical & Biological Engineering & Computing, PlosOne

Richard Kulpa, Human Movement Science, 2017, Sport Medicine (SMOA), 2017, CGVCVIP, 2017 Marc Christie, IEEE TVCG, The Visual Computer, CGF

### 10.1.4. Invited Talks

Anne-Hélène Olivier, VenLab, Brown University, Providence USA, August 2017

Anne-Hélène Olivier, IFSTTAR Marne la Vallée, November 2017

Antonio Mucherino, Institute of Computer Technology, TU Vienna, Austria. Invited by N. TaheriNejad. April 2017.

Antonio Mucherino, Research Center for Applied Sciences, Academia Sinica, Taipei, Taiwan. Invited by J-H. Lin. June 2017.

### 10.1.5. Scientific Expertise

Franck Multon, ANR expert, member of the ANR CPDS 4 "Santé Bien-être" through the Allistene national Alliance to design the next ANR call for projects, CRSNG (Canada)

Armel Crétual, 5 projects evaluated for UEFISCDI (The Executive Agency for Higher Education, Research, Development and Innovation Funding, Romania)

Marc Christie, expert for CIR (credit impot recherche - two cases in 2017), expert for ANR (CE 27)

#### 10.1.6. Research Administration

Franck Multon is member of the University Rennes2 Research steering committee "commission recherche", and Academic Council "CAC",

Franck Multon is member of the M2S Lab steering committee,

Franck Multon is member of the UFR-APS steering committee in University Rennes2

Ludovic Hoyet is an elected member of the Board of Managers for the French Computer Graphics Association (Association Française d'Informatique Graphique) since Oct. 2017

Ludovic Hoyet participated in the local MESR PhD Grant Auditions 2017

Georges Dumont is president of the elected group at scientific council of École Normale Supérieure de Rennes, member of the scientific council of École Normale Supérieure de Rennes

Georges Dumont is scientific head of Immerstar platforms (Immersia + Immermove) jointly for Inria and Irisa Partners

Richard Kulpa is member of the University Rennes2 Research steering committee "commission recherche", Academic Council "CAC" and Committee of International Affairs "CAI"il "CAC", Parait Bideau is director of the M2S Lab

Benoit Bideau is director of the M2S Lab

## **10.2. Teaching - Supervision - Juries**

### 10.2.1. Teaching

Doctorat : Ludovic Hoyet & Anne-Hélène Olivier, Evaluations en informatique graphique et réalité virtuelle : concepts généraux et cas pratiques, 2.25h, Journée Jeunes Chercheurs du GDR IG-RV 2017, France

Master : Franck Multon, co-leader of the IEAP Master (1 and 2) "Ingénierie et Ergonomie de l'Activité Physique", STAPS, University Rennes2, France

Master : Franck Multon, "Santé et Performance au Travail : étude de cas", leader of the module, 30H, Master 1 M2S, University Rennes2, France

Master : Franck Multon, "Analyse Biomécanique de la Performance Motrice", leader of the module, 30H, Master 1 M2S, University Rennes2, France

Master: Charles Pontonnier, "Numerical methods", leader of the module, 36H, Mechanics, École Spéciale Militaire de Saint-Cyr Coëtquidan, France

Master: Charles Pontonnier, "Numerical simulation of mechanical systems", leader of the module, 18H, Mechanics, École Spéciale Militaire de Saint-Cyr Coëtquidan, France

Master: Charles Pontonnier, "Analytical Mechanics", 40H, Mechanics, École Spéciale Militaire de Saint-Cyr Coëtquidan, France

Master: Charles Pontonnier, "Robotics: practical views", 24H, International cadets seminar, École Spéciale Militaire de Saint-Cyr Coëtquidan, France

Master: Charles Pontonnier, "Design and control of legged robots", leader of the module, 38H, Electronics, École Spéciale Militaire de Saint-Cyr Coëtquidan, France

Master: Charles Pontonnier, "Design, simulation and control of mechanical systems", leader of the module, 24H, Lecturers training in mechatronics, École Normale Supérieure de Rennes, France

Master: Charles Pontonnier, "Musculoskeletal modeling and ergonomics", 3H, Master 1 M2S, University Rennes2, France

Master : Georges Dumont, Responsible of the second year of the master Mechatronics, Rennes 1 University and École Normale Supérieure de Rennes, France

Master : Georges Dumont, Mechanical simulation in Virtual reality, 36H, Master Mechatronics, Rennes 1 University and École Normale Supérieure de Rennes, France

Master : Georges Dumont, Mechanics of deformable systems, 40H, Master FE, École Normale Supérieure de Rennes, France

Master : Georges Dumont, oral preparation to agregation competitive exam, 20H, Master FE, École Normale Supérieure de Rennes, France

Master : Georges Dumont, Vibrations in Mechanics, 10H, Master FE, École Normale Supérieure de Rennes, France

Master : Georges Dumont, Multibody Dynamics, 9H, Master FE, École Normale Supérieure de Rennes, France

Master : Georges Dumont, Finite Element method, 12H, Master FE, École Normale Supérieure de Rennes, France

Master : Ludovic Hoyet, Motion for Animation and Robotics, 9h, University Rennes 1, France

Master : Ludovic Hoyet, Motion Analysis and Gesture Recognition, 12h, INSA Rennes, France

Master : Anne-Hélène Olivier, "Biostatstiques", 7H, Master 2 APPCM, University Rennes2, France

Master : Anne-Hélène Olivier, "Evaluation fonctionnelle des pathologies motrices", 9H Master 2 APPCM, University Rennes2, France

Master : Anne-Hélène Olivier, "Maladie neurodégénératives : aspects biomécaniques", 2H Master 1 APPCM, University Rennes2, France

Master : Anne-Hélène Olivier, "Biostatstiques", 7H, Master 1 EOPS, University Rennes2, France

Master : Anne-Hélène Olivier, "Méthodologie", 4H, Master 1 EOPS/APPCM, University Rennes2, France

Master : Anne-Hélène Olivier, "Contrôle moteur : Boucle perceptivo-motrice", 3H, Master 1IEAP, Université Rennes 2, France

Master: Antonio Mucherino, "Programmation Parallèle", M1 en Informatique, 16H, University of Rennes, France

Master: Fabrice Lamarche, "Compilation pour l'image numérique", 29h, Master 1, ESIR, University of Rennes 1, France

Master: Fabrice Lamarche, "Synthèse d'images", 12h, Master 1, ESIR, University of Rennes 1, France

Master: Fabrice Lamarche, "Synthèse d'images avancée", 28h, Master 1, ESIR, University of Rennes 1, France

Master: Fabrice Lamarche, "Modélisation Animation Rendu", 36h, Master 2, ISTIC, University of Rennes 1, France

Master: Fabrice Lamarche, "Jeux vidéo", 26h, Master 2, ESIR, University of Rennes 1, France

Master: Fabrice Lamarche, "Motion for Animation and Robotics", 9h, Master 2 SIF, ISTIC, University of Rennes 1, France.

Master : Armel Crétual, "Méthodologie", leader of the module, 20H, Master 1 M2S, University Rennes2, France

Master : Armel Crétual, "Biostatstiques", leader of the module, 30H, Master 2 M2S, University Rennes2, France

Master : Richard Kulpa, "Boucle analyse-modélisation-simulation du mouvement", 27h, leader of the module, Master 2, Université Rennes 2, France

Master : Richard Kulpa, "Méthodes numériques d'analyse du geste", 27h, leader of the module, Master 2, Université Rennes 2, France

Master : Richard Kulpa, "Cinématique inverse", 3h, leader of the module, Master 2, Université Rennes 2, France

Master: Marc Christie, "Multimedia Mobile", Master 2, leader of the module, 32h, Computer Science, University of Rennes 1, France

Master: Marc Christie, "Projet Industriel Transverse", Master 2, 32h, leader of the module, Computer Science, University of Rennes 1, France

Master: Marc Christie, "Outils pour la Conception d'IHM", Master 2, 32h, leader of the module, Computer Science, University of Rennes 1, France

Licence : Franck Multon, "Ergonomie du poste de travail", Licence STAPS L2 & L3, University Rennes2, France

Licence: Charles Pontonnier, "Numerical control", leader of the module, Electronics, École Inter-Armes de Saint-Cyr Coëtquidan, France

Licence : Anne-Hélène Olivier, "Analyse cinématique du mouvement", 100H , Licence 1, University Rennes 2, France

Licence : Anne-Hélène Olivier, "Anatomie fonctionnelle", 7H , Licence 1, University Rennes 2, France

Licence : Anne-Hélène Olivier, "Effort et efficience", 12H , Licence 2, University Rennes 2, France

Licence : Anne-Hélène Olivier, "Locomotion et handicap", 12H , Licence 3, University Rennes 2, France

Licence : Anne-Hélène Olivier, "Biomécanique spécifique aux APA", 8H , Licence 3, University Rennes 2, France

Licence : Anne-Hélène Olivier, "Biomécanique du viellissement", 12H , Licence 3, University Rennes 2, France

Licence: Antonio Mucherino, "Informatique 1", 80H, L1, University of Rennes, France

Licence: Fabrice Lamarche, "Initiation à l'algorithmique et à la programmation", 56h, License 3, ESIR, University of Rennes 1, France

License: Fabrice Lamarche, "Programmation en C++", 46h, License 3, ESIR, University of Rennes 1, France

Licence: Fabrice Lamarche, "IMA", 24h, License 3, ENS Rennes, ISTIC, University of Rennes 1, France

Licence : Armel Crétual, "Analyse cinématique du mouvement", 100H, Licence 1, University Rennes 2, France

Licence : Richard Kulpa, "Biomécanique (dynamique en translation et rotation)", 48h, Licence 2, Université Rennes 2, France

Licence : Richard Kulpa, "Méthodes numériques d'analyse du geste", 48h, Licence 3, Université Rennes 2, France

Licence : Richard Kulpa, "Statistiques et informatique", 15h, Licence 3, Université Rennes 2, France

Licence : Marc Christie, "Programmation Impérative 1", leader of the module, University of Rennes 1, France

#### 10.2.2. Supervision

PhD: Antoine Muller, Contributions méthodologiques à l'analyse musculo-squelettique de l'humain dans l'objectif d'un compromis précision performance (Methodological contributions to the musculoskeletal analysis of humans with a trade-off between performance and precision), Ecole normale supérieure, Georges Dumont & Charles Pontonnier

PhD: Marion Morel, Suivi et étude des interactions pour l'analyse des tactiques durant un match de basket-ball, UPMC - University Rennes 2, septembre 2014, Catherine Achard & Séverine Dubuisson & Richard Kulpa

PhD in progress (beginning september 2017): Simon Hilt, Haptique Biofidèle pour l'Interaction en réalité virtuelle, Ecole normale supérieure, Georges Dumont& Charles Pontonnier

PhD in progress (beginning september 2017): Pierre Puchaud, Développement d'un modèle musculo-squelettique générique du soldat en vue du support de son activité physique, Ecole normale supérieure, Charles Pontonnier& Nicolas Bideau & Georges Dumont

PhD in progress: Rebecca Fribourg, Enhancing Avatars in Virtual Reality through Control, Interactions and Feedback, Sept. 2017, Ferran Argelaguet & Ludovic Hoyet & Anatole Lécuyer

PhD in progress: Florian Berton, Design of a virtual reality platform for studying immersion and behaviours in aggressive crowds, Nov. 2017, Ludovic Hoyet & Anne-Hélène Olivier & Julien Petré

PhD in progress: Sean D. Lynch, Perception visuelle du mouvement humain dans les interactions lors de tâches locomotrices, M2S - University Rennes 2, september 2015, Anne-Hélène Olivier & Richard Kulpa

PhD in progress: Amaury Louarn, A topology-driven approach to retargeting of filmic style in 3D environments, University of Rennes 1, Oct. 2017, Marc Christie & Fabrice Lamarche & Franck Multon

PhD in progress: Florence Gaillard, Evaluation en situation écologique des capacités fonctionnelles des membres supérieurs d'enfants hémiplégiques, University Rennes 2, decembre 2015, Armel Crétual & Isabelle Bonan

PhD in progress: Karim Jamal, Les effets des stimulations sensorielles par vibration sur les perturbations posturales secondaires à des troubles de la cognition spatiale après un Accident vasculaire Cérébrale, University Rennes 2, septembre 2016, Isabelle Bonan & Armel Crétual

PhD in progress: Lyse Leclercq, Intérêt dans les activités physiques du rétablissement de la fonction inertielle des membres supérieurs en cas d'amputation ou d'atrophie, University Rennes 2, septembre 2017, Armel Crétual

PhD in progress: Sean D. Lynch, Perception visuelle du mouvement humain dans les interactions lors de tâches locomotrices, M2S - University Rennes 2, septembre 2015, Anne-Hélène Olivier & Richard Kulpa

PhD in progress: Pierre Touzard, Suivi longitudinal du service de jeunes joueurs de tennis élite : identification biomécanique des facteurs de performance et de risque de blessures, University Rennes 2, septembre 2014, Benoit Bideau & Richard Kulpa & Caroline Martin

PhD in progress: Yacine Said Bouhalia, Approche transversale pour l'analyse et la reconnaissance de gestes 2D et 3D, INSA of Rennes, septembre 2015, Richard Kulpa & Franck Multon & Eric Anquetil

PhD in progress: Charles Faure, Stratégies coopératives et compétitives dans des tâches d'interactions physiques multiples, Université Rennes 2 - ENS Rennes, septembre 2016, Benoit Bideau & Richard Kulpa

PhD in progress: Théo Perrin, Evaluation de l'influence des feedbacks sur la capacité d'apprentissage dans le cadre d'interactions complexes entre joueurs et influence de ces feedbacks en fonction de l'activité sportive, Université Rennes 2 - ENS Rennes, septembre 2017, Benoit Bideau & Richard Kulpa

#### 10.2.3. Juries

PhD: Alexandra PIMENTA DOS SANTOS, Dynamical synthesis and analysis of healthy and pathological human walking, Université Pierre et Marie Curie, November 2017, Franck Multon, President

PhD: Galo Xavier MALDONADO TORO, Analysis and Generation of Highly Dynamic Motions of Anthropomorphic Systems: Application to Parkour. Université Fédérale Toulouse Midi-Pyrénées. November 2017, Franck Multon, Rapporteur.

PhD: 2017, Channarong Trakunsaranakom, Proposal for tangible, intuitive and collaborative design of manufactured products through virtual and augmented reality environments, Université de Grenoble-Alpes. 2017 june 21st, Georges Dumont Rapporteur.

PhD: 2017, Jingtao Chen, Biomechanical analysis of different aspects in virtual reality, Université de Grenoble-Alpes, 2017 january 30, Georges Dumont, Rapporteur.

PhD: 2017, Anne-Laure Kervellec, Introduction d'un logiciel dans la rééducation : optimisation et évaluation de l'engagement, Université Rennes 2, January 5th 2017, Armel Crétual, Examinateur

HDR: Sophie SAKKA, Entre l'humain et l'humanoïde. Ecole Centrale de Nantes. December 2017, Franck Multon, President.

## **10.3.** Popularization

Ludovic Hoyet, Journées de la Science 2017. Présentation et démo "Ma main à 6 doigts: Avatar et Réalité Virtuelle" à des classes de primaire et lycée.

Ludovic Hoyet, démonstrations pour les 50 ans d'Inria à Paris, "Ma main à 6 doigts: Avatar et Réalité Virtuelle".

# **11. Bibliography**

## **Publications of the year**

### **Doctoral Dissertations and Habilitation Theses**

[1] A. MULLER. *Methodological contributions to the human musculoskeletal simulation - performance and accuracy tradeoff*, École normale supérieure de Rennes, June 2017, https://tel.archives-ouvertes.fr/tel-01591365.

#### **Articles in International Peer-Reviewed Journal**

- [2] E. AUVINET, F. MULTON, V. MANNING, J. MEUNIER, J. P. COBB. Validity and sensitivity of the Longitudinal Asymmetry Index to detect gait asymmetry using Microsoft Kinect data, in "Gait and Posture", 2017, vol. 51, 17 [DOI: 10.1016/J.GAITPOST.2016.08.022], https://hal.inria.fr/hal-01359845.
- [3] A. L. CRUZ RUIZ, C. PONTONNIER, G. DUMONT.Low-Dimensional Motor Control Representations in *Throwing Motions*, in "Applied Bionics and Biomechanics", 2017, p. 1-26, https://hal.inria.fr/hal-01626143.
- [4] D. S. GONÇALVES, A. MUCHERINO, C. LAVOR, L. LIBERTI. Recent Advances on the Interval Distance Geometry, in "Journal of Global Optimization", January 2017, https://hal.inria.fr/hal-01429540.
- [5] D. HAERING, C. PONTONNIER, G. DUMONT. Which mathematical model best fit the maximal isometric torqueangle relationship of the elbow?, in "Computer Methods in Biomechanics and Biomedical Engineering", 2017, vol. 20, n<sup>o</sup> supp. 1, 2 [DOI : 10.1080/10255842.2017.1382881], https://hal.archives-ouvertes.fr/ hal-01533833.
- [6] M. HOFFMANN, D. HAERING, M. BEGON. Comparison between line and surface mesh models to represent the rotator cuff muscle geometry in musculoskeletal models, in "Computer Methods in Biomechanics and Biomedical Engineering", 2017, vol. 20, n<sup>o</sup> 11, p. 1175–1181 [DOI: 10.1080/10255842.2017.1340463], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01585975.
- [7] S. LYNCH, R. KULPA, L. MEERHOFF, J. PETTRÉ, A. CRÉTUAL, A.-H. OLIVIER. Collision avoidance behavior between walkers: global and local motion cues, in "IEEE Transactions on Visualization and Computer Graphics", 2017, vol. 23, p. 1 - 10 [DOI : 10.1109/TVCG.2017.2718514], https://hal.inria. fr/hal-01557763.

- [8] M. MOREL, C. ACHARD, R. KULPA, S. DUBUISSON. Automatic evaluation of sports motion: A generic computation of spatial and temporal errors, in "Image and Vision Computing", 2017, vol. 64, p. 67-78 [DOI: 10.1016/J.IMAVIS.2017.05.008], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01586401.
- [9] M. MOREL, C. ACHARD, R. KULPA, S. DUBUISSON.*Time-series averaging using constrained dynamic time warping with tolerance*, in "Pattern Recognition", 2018, vol. 74, p. 77-89 [DOI: 10.1016/J.PATCOG.2017.08.015], http://hal.upmc.fr/hal-01630288.
- [10] A. MULLER, C. PONTONNIER, G. DUMONT. *Uncertainty propagation in multibody human model dynamics*, in "Multibody System Dynamics", 2017, https://hal.inria.fr/hal-01462185.
- [11] A.-H. OLIVIER, J. BRUNEAU, R. KULPA, J. PETTRÉ. Walking with virtual people: Evaluation of locomotion interfaces in dynamic environments, in "IEEE Transactions on Visualization and Computer Graphics", 2017 [DOI: 10.1109/TVCG.2017.2714665], https://hal.inria.fr/hal-01557761.
- [12] P. PLANTARD, A. MULLER, C. PONTONNIER, G. DUMONT, H. P. H. SHUM, F. MULTON. Inverse dynamics based on occlusion-resistant Kinect data: Is it usable for ergonomics?, in "International Journal of Industrial Ergonomics", 2017, vol. 61, p. 71-80 [DOI : 10.1016/J.ERGON.2017.05.010], https://hal.inria.fr/hal-01527325.
- [13] P. PLANTARD, H. P. H. SHUM, A.-S. LE PIERRES, F. MULTON. Validation of an ergonomic assessment method using Kinect data in real workplace conditions, in "Applied Ergonomics", 2017, vol. 65, p. 562-569 [DOI: 10.1016/J.APERGO.2016.10.015], https://hal.inria.fr/hal-01393066.
- [14] P. PLANTARD, H. SHUM, F. MULTON. Filtered Pose Graph for Efficient Kinect Pose Reconstruction, in "Multimedia Tools and Applications", 2017, vol. 76, n<sup>o</sup> 3, 24 [DOI: 10.1007/s11042-016-3546-4], https:// hal.inria.fr/hal-01332711.
- [15] P. PLANTARD, H. SHUM, F. MULTON. Usability of corrected Kinect measurement for ergonomic evaluation in constrained environment, in "International Journal of Human Factors Modelling and Simulation", 2017, vol. 5, n<sup>o</sup> 4, p. 338 - 353 [DOI: 10.1504/IJHFMS.2017.087018], https://hal.inria.fr/hal-01612939.
- [16] P. PLANTARD, A. SOREL, N. BIDEAU, C. PONTONNIER. Motion Adaptation in Fencing Lunges: a Pilot Study, in "Computer Methods in Biomechanics and Biomedical Engineering", 2017, vol. 20, p. 161-162 [DOI: 10.1080/10255842.2017.1382912], https://hal.inria.fr/hal-01534895.
- [17] C. VASSALLO, A.-H. OLIVIER, P. SOUÈRES, A. CRÉTUAL, O. STASSE, J. PETTRÉ. How do walkers avoid a mobile robot crossing their way?, in "Gait and Posture", January 2017, vol. 51, p. 97-103, https://arxiv.org/ abs/1609.07955 [DOI: 10.1016/J.GAITPOST.2016.09.022], https://hal.archives-ouvertes.fr/hal-01371202.

#### **International Conferences with Proceedings**

- [18] A. BERNARDIN, L. HOYET, A. MUCHERINO, D. S. GONÇALVES, F. MULTON. Normalized Euclidean Distance Matrices for Human Motion Retargeting, in "MIG 2017 - Motion in Games", Barcelona, Spain, November 2017 [DOI: 10.1145/3136457.3136466], https://hal.inria.fr/hal-01632850.
- [19] D. HAERING, C. PONTONNIER, N. BIDEAU, G. NICOLAS, G. DUMONT. Task Specific Maximal Elbow Torque Model For Ergonomic Evaluation, in "XXVI Congress of the International Society of Biomechanics", Brisbane, Australia, July 2017, https://hal.inria.fr/hal-01495065.

- [20] S. HILT, C. PONTONNIER, G. DUMONT. Model Based Compensation for Low Mass Objects Haptic Manipulation in Virtual Environments, in "Proceedings of EuroVR conference 2017", Laval, France, Lecture notes in computer sciences, December 2017, p. 1-16, https://hal.inria.fr/hal-01626147.
- [21] L. MEERHOFF, J. PETTRÉ, R. KULPA, S. LYNCH, A. CRÉTUAL, A.-H. OLIVIER. Simultaneous and sequential affordances of collision avoidance between multiple pedestrians, in "ICPA 2017 - 19th biannual International Conference on Perception and Action", Séoul, South Korea, July 2017, https://hal.inria.fr/hal-01646507.
- [22] A. MUCHERINO, D. S. GONÇALVES, A. BERNARDIN, L. HOYET, F. MULTON. A Distance-Based Approach for Human Posture Simulations, in "FedCSIS 2017 - Federated Conference on Computer Science and Information Systems", Prague, Czech Republic, 2017 Federated Conference on Computer Science and Information Systems (FedCSIS), IEEE, September 2017, vol. 11, p. 441-444 [DOI : 10.15439/2017F98], https://hal.inria.fr/hal-01632860.
- [23] A. MUCHERINO, D. S. GONÇALVES. An Approach to Dynamical Distance Geometry, in "GSI 2017 -International Conference on Geometric Science of Information", Paris, France, GSI 2017: Geometric Science of Information, Springer, November 2017, vol. 10589, p. 821-829 [DOI: 10.1007/978-3-319-68445-1\_94], https://hal.inria.fr/hal-01632847.
- [24] A. MULLER, M. CHAUWIN, C. PONTONNIER, G. DUMONT. Influence of Physiological Constraints on a Subject-specific BSIP Calibration, in "XXVI Congress of the International Society of Biomechanics", Brisbane, Australia, International Society of Biomechanics, July 2017, https://hal.inria.fr/hal-01494298.
- [25] A. MULLER, F. DEMORE, C. PONTONNIER, G. DUMONT.*MusIC MAKES THE MUSCLES WORK TO-GETHER*, in "XVI International Symposium on Computer Simulation in Biomechanics", Brisbane, Australia, July 2017, 2, https://hal.inria.fr/hal-01502184.
- [26] A.-H. OLIVIER, J. BRUNEAU, R. KULPA, J. PETTRÉ. Walking with virtual people: Evaluation of locomotion interfaces in dynamic environments, in "IEEE VR 2017 - IEEE Virtual Reality", Los Angeles, United States, IEEE, March 2017, vol. PP, n<sup>o</sup> 99, p. 1-13 [DOI: 10.1109/TVCG.2017.2714665], https://hal.inria.fr/hal-01661211.
- [27] J. OMER, A. MUCHERINO.Integer Programming for the Search of a Discretization Order in Distance Geometry Problems, in "ROADEF 2017 - 18ème congrès annuel de la Société Française de Recherche Opérationnelle et d'Aide à la Décision", Metz, France, February 2017, p. 1-2, https://hal.inria.fr/hal-01632864.
- [28] P. TOUZARD, R. KULPA, M. CAROLINE, B. BIDEAU. Biomechanical analysis of tennis serve in young elite players: effect of the "waiter's serve" on upper limb loads, in "European College of Sport Science", Ruhr Metropolis, Germany, January 2017, https://hal.inria.fr/hal-01676767.
- [29] H.-Y. WU, Q. GALVANE, C. LINO, M. CHRISTIE. Analyzing Elements of Style in Annotated Film Clips, in "WICED 2017 - Eurographics Workshop on Intelligent Cinematography and Editing", Lyon, France, April 2017, 7 [DOI: 10.2312/WICED.20171068], https://hal.inria.fr/hal-01492523.

#### **National Conferences with Proceeding**

[30] A. MULLER, D. HAERING, C. PONTONNIER, G. DUMONT. Non-invasive techniques for musculoskeletal model calibration, in "Congrès Français de Mécanique", Lille, France, August 2017, https://hal.inria.fr/hal-01524330.

### **Conferences without Proceedings**

- [31] S. Y. BOULAHIA, E. ANQUETIL, F. MULTON, R. KULPA. Dynamic hand gesture recognition based on 3D pattern assembled trajectories, in "IPTA 2017 - 7th IEEE International Conference on Image Processing Theory, Tools and Applications", Montreal, Canada, November 2017, p. 1-6, https://hal.archives-ouvertes.fr/ hal-01666377.
- [32] S. Y. BOULAHIA, E. ANQUETIL, R. KULPA, F. MULTON.3D Multistroke Mapping (3DMM): Transfer of Hand-Drawn Pattern Representation for Skeleton-Based Gesture Recognition, in "12th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2017)", Washington, DC., United States, May 2017, p. 462 - 467 [DOI: 10.1109/FG.2017.63], https://hal.archives-ouvertes.fr/hal-01555452.
- [33] S. LYNCH, J. PETTRÉ, J. BRUNEAU, R. KULPA, A. CRÉTUAL, A.-H. OLIVIER. Effect of mutual gaze on collision avoidance behavior during human walking, in "AFRV", Rennes, France, October 2017, https://hal. archives-ouvertes.fr/hal-01645801.
- [34] H.-Y. WU, T.-Y. LI, M. CHRISTIE. Using ECPs for Interactive Applications in Virtual Cinematography, in "Eurographics Workshop on Intelligent Cinematography and Editing", Lyon, France, April 2017, https://hal.archives-ouvertes.fr/hal-01571418.

### **Scientific Popularization**

[35] L. HOYET, J. PETTRÉ. La simulation de foule au service du divertissement, in "Interstices", April 2017, https:// hal.inria.fr/hal-01533687.

#### **Other Publications**

[36] S. LYNCH, R. KULPA, L. MEERHOFF, J. PETTRÉ, A. CRÉTUAL, A.-H. OLIVIER. Effect of local limb cues in the prediction of global motion during collision avoidance, October 2017, ACAPS, Poster, https://hal.archivesouvertes.fr/hal-01645924.
# **Project-Team MYRIADS**

# Design and Implementation of Autonomous Distributed Systems

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: CNRS Institut national des sciences appliquées de Rennes Université Rennes 1

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Distributed Systems and middleware

# **Table of contents**

2. Overall Objectives     693       2.1. General Objectives     693       2.2. Context     693       2.3. Challenges     694       3.4. Research Program     695       3.1. Introduction     695       3.2. Scaling clouds     695       3.2. Resource management in mobile edge clouds     695       3.2.1. Resource management in mobile edge clouds     696       3.3. Greening clouds     696       3.3. Greening clouds     696       3.3. I. Smart grids and clouds     696       3.3. Energy cost models     697       3.3. Energy cost models     697       3.4. Security monitoring SLO     697       3.4. Security monitoring SLO     697       3.4. Security monitoring SLO     699       3.5. Experimenting with Clouds     699       3.5.1. Experimentation methodologies for clouds     699       3.5.2. Use cases     699       3.5.2. Use cases     699       3.5.2. Use cases     700       6. New Software and Platforms     700       6. SinGrid     702       6. SinGrid     702       7. New Results     703       7	1.	Personnel	692
2.1.     Genéral Objectives     693       2.2.     Context     693       2.3.     Challenges     694       3. Research Program     695       3.1.     Introduction     695       3.2.     Resource management in hierarchical clouds     695       3.2.     Resource management in mobile edge clouds     695       3.2.     Resource management in mobile edge clouds     696       3.3.     Greening clouds     696       3.3.     Start protection in multi-cloud environments     696       3.3.     Greening clouds     697       3.4.     Security monitoring SLO     697       3.4.     Security monitoring SLO     697       3.4.1.     Security monitoring SLO     698       3.5.1.     Experimenting with Clouds     698       3.5.1.     Experimenting with Clouds     698       3.5.1.     Experimenting with Clouds     699       3.5.2.     Use cases     699       4.     Application Domains     699       5.     Highlights of the Year     700       6.1.     ConPaaS     700       6.2.	2.	Overall Objectives	693
2.2. Context     693       2.3. Challenges     694       3. Research Program     695       3.1. Introduction     695       3.2. Scaling clouds     695       3.2.1. Resource management in biele dge clouds     695       3.2.2. Resource management in mobile edge clouds     695       3.2.3. Self-optimizing applications in multi-cloud environments     696       3.3. Greening clouds     696       3.3. Is smart grids and clouds     697       3.3. Energy-aware users     697       3.4. Security monitoring SLO     697       3.4. Security monitoring SLO     697       3.4. Security monitoring SLO     697       3.5. Experimenting with Clouds     698       3.5. Lexperimenting with Clouds     699       3.5.1. Experimenting with Clouds     699       3.5.2. Use cases     699       4. Application Domains     700       6. New Software and Platforms     700       6.1. ConPaaS     700       7.0. New Results     703       7.1. Scaling Clouds     703       7.1. Scaling Clouds     703       7.1. Scaling Clouds     703       7.1. Scaling Clouds		2.1. General Objectives	693
2.3. Challenges     694       3. Research Program     695       3.1. Introduction     695       3.2. Scaling clouds     695       3.2.1. Resource management in hierarchical clouds     695       3.2.2. Resource management in mobile edge clouds     695       3.2.3. Self-optimizing applications in multi-cloud environments     696       3.3.3. Serence yaware users     697       3.4.1. Security monitoring SLO     697       3.5.1. Experimentation methodologies for clouds     698       3.5.1. Experimentation methodologies for clouds     699       3.6.1. ConPaaS     700       6.1. ConPaaS     700       6.2. GinFlow     701       6.3. Merkat     701       7.4.4. PaaSage Adapter     701       6.5. SinGrid     702       7.1. Scaling Clouds     703       7.1.		2.2. Context	693
3. Research Program     695       3.1. Introduction     695       3.2. Scaling clouds     695       3.2.1. Resource management in hierarchical clouds     695       3.2.2. Resource management in mobile edge clouds     695       3.2.3. Self-optimizing applications in multi-cloud environments     696       3.3. Greening clouds     696       3.3.1. Energy cost models     697       3.3.2. Energy cost models     697       3.3.3. Energy-aware users     697       3.4. Securing clouds     697       3.4. Securing monitoring SLO     697       3.4. Securing monitoring SLO     697       3.4. Experimenting with Clouds     698       3.5. Experimentation methodologies for clouds     699       3.5.1. Experimentation methodologies for clouds     699       3.5.2. Use cases     699       9.9. Highlights of the Year     700       6. New Software and Platforms     700       6.1. ConPaaS     702       6.2. GinFlow     701       6.3. Merkat     701       6.4. SimGrid     702       7.1. Scaling Clouds     703       7.1. Scaling Clouds     703		2.3. Challenges	694
3.1.     Introduction     695       3.2.     Scaling clouds     695       3.2.1.     Resource management in mobile edge clouds     695       3.2.2.     Resource management in mobile edge clouds     695       3.2.3.     Self-optimizing applications in multi-cloud environments     696       3.3.1.     Smart grids and clouds     696       3.3.1.     Smart grids and clouds     697       3.3.2.     Energy cost models     697       3.3.3.     Energy aware users     697       3.4.     Security monitoring SLO     697       3.4.     Security monitoring SLO     697       3.5.1.     Experimentation methodologies for clouds     698       3.5.1.     Experimentation methodologies for clouds     699       3.5.1.     Experimentation methodologies for clouds     699       5.5.     Use cases     699       5.6.     Use cases     700       6.     New Software and Platforms     700       6.     SinGrid     702       6.     SinGrid     702       6.     SinGrid     702       6.     SinGrid     703 <td>3.</td> <td>Research Program</td> <td> 695</td>	3.	Research Program	695
3.2.     Scaling clouds     695       3.2.1.     Resource management in hierarchical clouds     695       3.2.2.     Resource management in mobile edge clouds     695       3.2.3.     Self-optimizing applications in multi-cloud environments     696       3.3.     Greening clouds     696       3.3.     Energy cost models     697       3.4.     Securing clouds     697       3.4.     Securing clouds     697       3.4.1.     Security monitoring SLO     697       3.4.2.     Data Protection in Cloud-based IoT Services     698       3.5.1.     Experimenting with Clouds     699       3.5.2.     Use cases     699       4. Application Domains     699     700       6. New Software and Platforms     700       6. New Software and Platforms     700       6.3.     GinFlow     701       6.5.     SAIDS     702       7.1.     Scaling Clouds     703       7.1.     Scaling Clouds     703       7.1.     Scaling Clouds     703       7.1.     Scaling Clouds     703       7.1.     Scal		3.1. Introduction	695
3.2.1.     Resource management in hierarchical clouds     695       3.2.2.     Resource management in mobile edge clouds     695       3.2.3.     Self-optimizing applications in multi-cloud environments     696       3.3.     Greening clouds     696       3.3.     Energy cost models     697       3.3.     Energy cost models     697       3.3.     Energy cost models     697       3.4.     Securing clouds     697       3.4.     Securiny monitoring SLO     697       3.4.     Securiny monitoring SLO     697       3.4.1.     Experimenting with Clouds     698       3.5.1.     Experimentation methodologies for clouds     699       3.5.2.     Use cases     699       4.     Application Domains     609       5.1.     Experimentation methodologies for clouds     699       6.2.     GinFlow     700       6.1.     ConPaaS     700       6.2.     GinFlow     701       6.3.     SAIDS     702       6.4.     PaaSage Adapter     701       6.5.     SAIDS     702       6.7		3.2. Scaling clouds	695
3.2.2.     Resource management in mobile edge clouds     695       3.2.3.     Self-optimizing applications in multi-cloud environments     696       3.3.     Greening clouds     696       3.3.1.     Smart grids and clouds     696       3.3.2.     Energy cost models     697       3.3.3.     Energy cost models     697       3.4.3.     Securing clouds     697       3.4.4.     Security monitoring SLO     697       3.4.1.     Security monitoring SLO     697       3.4.2.     Data Protection in Cloud-based IoT Services     698       3.5.1.     Experimenting with Clouds     698       3.5.1.     Experimenting with Clouds     698       3.5.2.     Use cases     699       4.     Application Domains     699       5.     Highlights of the Year     700       6.     New Software and Platforms     700       6.1.     ConPaaS     700       6.2.     GinFlow     701       6.3.     Merkat     701       6.4.     PaaSage Adapter     701       6.5.     SimGrid     702       7		3.2.1. Resource management in hierarchical clouds	695
3.2.3.     Self-optimizing applications in multi-cloud environments     696       3.3.     Greening clouds     696       3.3.1.     Smart grids and clouds     696       3.3.2.     Energy cost models     697       3.3.3.     Energy-aware users     697       3.4.     Securing clouds     697       3.4.1.     Securing clouds     697       3.4.2.     Data Protection in Cloud-based IoT Services     698       3.5.1.     Experimenting with Clouds     699       3.5.2.     Use cases     699       3.5.3.     Experimentation methodologies for clouds     699       3.5.4.     Experimentation methodologies for clouds     699       3.5.2.     Use cases     699       4.     Application Domains     699       5.5.     Highlights of the Year     700       6.1.     ConPaaS     700       6.2.     GinFlow     701       6.3.     Merkat     701       6.4.     PaaSage Adapter     702       6.5.     SinGrid     702       7.1.     Kedge Clouds     703       7.1.1. <td< td=""><td></td><td>3.2.2. Resource management in mobile edge clouds</td><td>695</td></td<>		3.2.2. Resource management in mobile edge clouds	695
3.3. Greening clouds     696       3.3.1. Smart grids and clouds     696       3.3.2. Energy-cost models     697       3.3.3. Energy-aware users     697       3.4. Securing clouds     697       3.4.1. Security monitoring SLO     697       3.4.2. Data Protection in Cloud-based IoT Services     698       3.5. Experimenting with Clouds     699       3.5.1. Experimentation methodologies for clouds     699       3.5.2. Use cases     699       4. Application Domains     699       5. Highlights of the Year     700       6. New Software and Platforms     700       6.1. ConPaaS     700       6.2. GinFlow     701       6.3. Merkat     701       6.4. PaaSage Adapter     701       6.5. SAIDS     702       6.6. SimGrid     702       7.1. Fog Computing     703       7.1.1. Fog Computing     703       7.1.2. Edge Cloud     703       7.1.3. Community Clouds     704       7.1.4. Evaluation of Data Stream Processing Frameworks in Clouds     704       7.1.5. Stream Processing for Maritime Surveillance     704       7.1.6. Adaptive deployment f		3.2.3. Self-optimizing applications in multi-cloud environments	696
3.3.1.     Smart grids and clouds     696       3.3.2.     Energy cost models     697       3.3.     Energy-aware users     697       3.4.     Securing clouds     697       3.4.     Security monitoring SLO     697       3.4.1.     Security monitoring SLO     697       3.4.2.     Data Protection in Cloud-based IoT Services     698       3.5.     Experimenting with Clouds     698       3.5.1.     Experimentation methodologies for clouds     699       3.5.2.     Use cases     699       3.5.2.     Use cases     699       3.5.4.     Key cases     699       3.5.5.     Highlights of the Year     700       6.     New Software and Platforms     700       6.1.     ConPaats     701       6.3.     Merkat     701       6.4.     PaaSage Adapter     701       6.5.     SAIDS     702       6.6.     SimGrid     702       7.1.     Fog Computing     703       7.1.1.     Fog Computing     703       7.1.2.     Edge Cloud     703 <tr< td=""><td></td><td>3.3. Greening clouds</td><td>696</td></tr<>		3.3. Greening clouds	696
3.3.2. Energy cost models     697       3.3.3. Energy-aware users     697       3.4. Securing clouds     697       3.4. Securing clouds     697       3.4.1. Security monitoring SLO     697       3.4.2. Data Protection in Cloud-based IoT Services     698       3.5. Experimentation methodologies for clouds     699       3.5.1. Experimentation methodologies for clouds     699       3.5.2. Use cases     699       4. Application Domains     700       6. New Software and Platforms     700       6.1. ConPaaS     700       6.2. GinFlow     701       6.3. Merkat     701       6.4. PaaSage Adapter     701       6.5. SAIDS     702       6.6. SimGrid     702       6.7. DiFFuSE     703       7.1. Scaling Clouds     703       7.1.1. Fog Computing     703       7.1.3. Community Clouds     704       7.1.4. Evaluation of Data Stream Processing Frameworks in Clouds     704       7.1.5. Stream Processing for Maritime Surveillance     704       7.1.6. Adaptive deployment for multi-cloud applications     705       7.1.7. Application configuration and reconfiguration in multi-eloud en		3.3.1. Smart grids and clouds	696
3.3.3.     Energy-aware users     697       3.4.     Security monitoring SLO     697       3.4.1.     Security monitoring SLO     697       3.4.2.     Data Protection in Cloud-based IoT Services     698       3.5.     Experimenting with Clouds     699       3.5.1.     Experimentation methodologies for clouds     699       3.5.2.     Use cases     699       4.     Application Domains     699       5.     Highlights of the Year     700       6.     New Software and Platforms     700       6.1.     ConPaaS     700       6.2.     GinFlow     701       6.3.     Merkat     701       6.4.     PaaSage Adapter     702       6.6.     SimGrid     702       6.6.     SimGrid     702       7.1.     Scaling Clouds     703       7.1.1.     Fog Computing     703       7.1.2.     Edge Cloud     703       7.1.3.     Community Clouds     704       7.1.4.     Evaluation of Data Stream Processing Frameworks in Clouds     704       7.1.5.     Stream Processing fo		3.3.2. Energy cost models	697
3.4.     Security monitoring SLO     697       3.4.1.     Security monitoring SLO     697       3.4.2.     Data Protection in Cloud-based IoT Services     698       3.5.     Experimenting with Clouds     698       3.5.1.     Experimentation methodologies for clouds     699       3.5.2.     Use cases     699       4.     Application Domains     699       5.     Highlights of the Year     700       6.     New Software and Platforms     700       6.1.     ConPaaS     700       6.2.     GinFlow     701       6.3.     Merkat     701       6.4.     PaaSage Adapter     701       6.5.     SAIDS     702       6.6.     SimGrid     702       7.1.     Scaling Clouds     703       7.1.1.     Fog Computing     703       7.1.2.     Edge Cloud     703       7.1.3.     Community Clouds     704       7.1.4.     Evaluation of Data Stream Processing Frameworks in Clouds     704       7.1.5.     Stream Processing for Maritime Surveillance     704       7.1.6.     <		3.3.3. Energy-aware users	697
3.4.1.     Security monitoring SLO     697       3.4.2.     Data Protection in Cloud-based IoT Services     698       3.5.     Experimentation methodologies for clouds     699       3.5.1.     Experimentation methodologies for clouds     699       3.5.2.     Use cases     699       4.     Application Domains     700       6.     New Software and Platforms     700       6.1.     ConPaaS     700       6.2.     GinFlow     701       6.3.     Merkat     701       6.4.     PaaSage Adapter     702       6.6.     SimGrid     702       6.7.     DiFFuSE     703       7.1.     Scaling Clouds     703       7.1.1.     Fog Computing     703       7.1.2.     Edge Cloud     703       7.1.3.     Community Clouds     704       7.1.4.     Evaluation of Data Stream Processing Frameworks in Clouds     704       7.1.5.     Stream Processing for Maritime Surveillance     704       7.1.6.     Adaptive resource management for high-performance, multi-sensor systems     705       7.1.7.     Application configuratio		3.4. Securing clouds	697
3.4.2. Data Protection in Cloud-based IoT Services     698       3.5. Experimenting with Clouds     698       3.5.1. Experimentation methodologies for clouds     699       3.5.2. Use cases     699       4. Application Domains     699       5. Highlights of the Year     700       6. New Software and Platforms     700       6.1. ConPaaS     700       6.2. GinFlow     701       6.3. Merkat     701       6.4. PaaSage Adapter     701       6.5. SAIDS     702       6.6. SimGrid     702       6.7. DiFFuSE     703       7.1. Scaling Clouds     703       7.1.1. Fog Computing     703       7.1.2. Edge Cloud     703       7.1.3. Community Clouds     704       7.1.4. Evaluation of Data Stream Processing Frameworks in Clouds     704       7.1.5. Stream Processing for Maritime Surveillance     704       7.1.6. Adaptive deployment for multi-cloud applications     705       7.1.7. Application configuration and reconfiguration in multi-cloud environments     705       7.1.8. Adaptive resource management for high-performance, multi-sensor systems     705       7.1.2. Edge Clouds     706 <t< td=""><td></td><td>3.4.1. Security monitoring SLO</td><td>697</td></t<>		3.4.1. Security monitoring SLO	697
3.5.     Experimenting with Clouds     698       3.5.1.     Experimentation methodologies for clouds     699       3.5.2.     Use cases     699       4.     Application Domains     699       5.     Highlights of the Year     700       6.     New Software and Platforms     700       6.1.     ConPaaS     700       6.2.     GinFlow     701       6.3.     Merkat     701       6.4.     PaaSage Adapter     702       6.5.     SAIDS     702       6.6.     SimGrid     702       6.7.     DiFFuSE     703       7.1.     Scaling Clouds     703       7.1.1.     Fog Computing     703       7.1.2.     Edge Cloud     703       7.1.3.     Community Clouds     704       7.1.4.     Evaluation of Data Stream Processing Frameworks in Clouds     704       7.1.5.     Stream Processing for Maritime Surveillance     704       7.1.6.     Adaptive deployment for multi-cloud applications     705       7.1.8.     Adaptive resource management for high-performance, multi-sensor systems     705		3.4.2. Data Protection in Cloud-based IoT Services	698
3.5.1.     Experimentation methodologies for clouds     699       3.5.2.     Use cases     699       4.     Application Domains     699       5.     Highlights of the Year     700       6.     New Software and Platforms     700       6.1.     ConPaaS     700       6.2.     GinFlow     701       6.3.     Merkat     701       6.4.     PaaSage Adapter     701       6.5.     SAIDS     702       6.6.     SimGrid     702       6.7.     DiFFuSE     703       7.1.     Fog Computing     703       7.1.1.     Fog Computing     703       7.1.2.     Edge Cloud     703       7.1.3.     Community Clouds     704       7.1.4.     Evaluation of Data Stream Processing Frameworks in Clouds     704       7.1.4.     Evaluation and reconfiguration in multi-cloud environments     705       7.1.6.     Adaptive resource management for high-performance, multi-sensor systems     705       7.1.8.     Adaptive resource management for high-performance, multi-sensor systems     705       7.2.1.     Energy Models		3.5. Experimenting with Clouds	698
3.5.2.     Use cases     699       4.     Application Domains     699       5.     Highlights of the Year     700       6.     New Software and Platforms     700       6.1.     ConPaaS     700       6.2.     GinFlow     701       6.3.     Merkat     701       6.4.     PaaSage Adapter     701       6.5.     SAIDS     702       6.6.     SimGrid     702       6.7.     DiFFuSE     703       7.1.     Scaling Clouds     703       7.1.1.     Fog Computing     703       7.1.2.     Edge Cloud     703       7.1.3.     Community Clouds     704       7.1.4.     Evaluation of Data Stream Processing Frameworks in Clouds     704       7.1.5.     Stream Processing for Maritime Surveillance     704       7.1.6.     Adaptive deployment for multi-cloud applications     705       7.1.8.     Adaptive resource management for high-performance, multi-sensor systems     706       7.2.1.     Energy Models     706       7.2.2.     Exploiting Renewable Energy in Clouds     707		3.5.1. Experimentation methodologies for clouds	699
4. Application Domains     699       5. Highlights of the Year     700       6. New Software and Platforms     700       6.1. ConPaaS     700       6.2. GinFlow     701       6.3. Merkat     701       6.4. PaaSage Adapter     701       6.5. SAIDS     702       6.6. SimGrid     702       6.7. DiFFuSE     703       7.1. Scaling Clouds     703       7.1.1. Fog Computing     703       7.1.2. Edge Cloud     703       7.1.3. Community Clouds     704       7.1.4. Evaluation of Data Stream Processing Frameworks in Clouds     704       7.1.5. Stream Processing for Maritime Surveillance     704       7.1.6. Adaptive deployment for multi-cloud applications     705       7.1.7. Application configuration and reconfiguration in multi-cloud environments     705       7.1.8. Adaptive resource management for high-performance, multi-sensor systems     706       7.2.1. Energy Models     706       7.2.2. Exploiting Renewable Energy in Clouds     707       7.3. Security Monitoring in Clouds     707       7.3. Security Monitoring in Clouds     707       7.3.2. Risk assessment in clouds     708  <		3.5.2. Use cases	699
5.     Highlights of the Year     700       6.     New Software and Platforms     700       6.1.     ConPaaS     700       6.2.     GinFlow     701       6.3.     Merkat     701       6.4.     PaaSage Adapter     701       6.5.     SAIDS     702       6.6.     SimGrid     702       6.7.     DiFFuSE     703       7.1.     Scaling Clouds     703       7.1.1.     Fog Computing     703       7.1.2.     Edge Cloud     703       7.1.3.     Community Clouds     704       7.1.4.     Evaluation of Data Stream Processing Frameworks in Clouds     704       7.1.5.     Stream Processing for Maritime Surveillance     704       7.1.6.     Adaptive deployment for multi-cloud applications     705       7.1.7.     Application configuration and reconfiguration in multi-cloud environments     705       7.2.     Greening Clouds     706       7.2.1.     Energy Models     706       7.2.2.     Exploiting Renewable Energy in Clouds     706       7.2.3.     Smart Grids     707	4.	Application Domains	699
6. New Software and Platforms     700       6.1. ConPaaS     700       6.2. GinFlow     701       6.3. Merkat     701       6.4. PaaSage Adapter     701       6.5. SAIDS     702       6.6. SimGrid     702       6.7. DiFFuSE     703       7. New Results     703       7.1. Scaling Clouds     703       7.1.1. Fog Computing     703       7.1.2. Edge Cloud     703       7.1.3. Community Clouds     704       7.1.4. Evaluation of Data Stream Processing Frameworks in Clouds     704       7.1.5. Stream Processing for Maritime Surveillance     704       7.1.6. Adaptive deployment for multi-cloud applications     705       7.1.7. Application configuration and reconfiguration in multi-cloud environments     705       7.1.8. Adaptive resource management for high-performance, multi-sensor systems     705       7.2. Greening Clouds     706       7.2.1. Energy Models     706       7.2.2. Exploiting Renewable Energy in Clouds     706       7.2.3. Smart Grids     707       7.3. Security Monitoring in Clouds     707       7.3.1. Security Monitoring in Clouds     707       7.3.2. Ri	5.	Highlights of the Year	700
6.1. ConPaas7006.2. GinFlow7016.3. Merkat7016.4. PaaSage Adapter7016.5. SAIDS7026.6. SimGrid7026.7. DiFFuSE7037. New Results7037.1. Scaling Clouds7037.1.1. Fog Computing7037.1.2. Edge Cloud7037.1.3. Community Clouds7047.1.4. Evaluation of Data Stream Processing Frameworks in Clouds7047.1.5. Stream Processing for Maritime Surveillance7047.1.6. Adaptive deployment for multi-cloud applications7057.1.7. Application configuration and reconfiguration in multi-sensor systems7057.1.8. Adaptive resource management for high-performance, multi-sensor systems7067.2.1. Energy Models7067.2.2. Exploiting Renewable Energy in Clouds7067.2.3. Smart Grids7077.3.1. Security Monitoring in Clouds7077.3.2. Risk assessment in clouds708	6.	New Software and Platforms	700
6.2. GinFlow7016.3. Merkat7016.4. PaaSage Adapter7016.5. SAIDS7026.6. SimGrid7026.7. DiFFuSE7037. New Results7037.1. Scaling Clouds7037.1.1. Fog Computing7037.1.2. Edge Cloud7037.1.3. Community Clouds7047.1.4. Evaluation of Data Stream Processing Frameworks in Clouds7047.1.5. Stream Processing for Maritime Surveillance7047.1.6. Adaptive deployment for multi-cloud applications7057.1.7. Application configuration and reconfiguration in multi-sensor systems7057.2. Greening Clouds7067.2.1. Energy Models7067.2.2. Exploiting Renewable Energy in Clouds7067.2.3. Smart Grids7077.3.1. Security Monitoring in Clouds7077.3.2. Risk assessment in clouds708		6.1. ConPaaS	700
6.3. Merkat7016.4. PaaSage Adapter7016.5. SAIDS7026.6. SimGrid7026.7. DiFFuSE7037. New Results7037.1. Scaling Clouds7037.1.1. Fog Computing7037.1.2. Edge Cloud7037.1.3. Community Clouds7047.1.4. Evaluation of Data Stream Processing Frameworks in Clouds7047.1.5. Stream Processing for Maritime Surveillance7047.1.6. Adaptive deployment for multi-cloud applications7057.1.7. Application configuration and reconfiguration in multi-cloud environments7057.1.8. Adaptive resource management for high-performance, multi-sensor systems7057.2. Greening Clouds7067.2.1. Energy Models7067.2.2. Exploiting Renewable Energy in Clouds7077.2.3. Smart Grids7077.3.1. Securing Clouds7077.3.1. Securing Clouds7077.3.2. Risk assessment in clouds708		6.2. GinFlow	701
6.4.PaaSage Adapter7016.5.SAIDS7026.6.SimGrid7026.7.DiFFuSE7037.New Results7037.1.Scaling Clouds7037.1.1.Fog Computing7037.1.2.Edge Cloud7037.1.3.Community Clouds7047.1.4.Evaluation of Data Stream Processing Frameworks in Clouds7047.1.5.Stream Processing for Maritime Surveillance7047.1.6.Adaptive deployment for multi-cloud applications7057.1.7.Application configuration and reconfiguration in multi-cloud environments7057.1.8.Adaptive resource management for high-performance, multi-sensor systems7067.2.1.Energy Models7067.2.2.Exploiting Renewable Energy in Clouds7077.2.4.Involving Users in Energy Saving7077.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds708		6.3. Merkat	701
6.3.SAIDS7026.6.SimGrid7026.7.DiFFuSE7037.1.Scaling Clouds7037.1.Fog Computing7037.1.1.Fog Computing7037.1.2.Edge Cloud7037.1.3.Community Clouds7047.1.4.Evaluation of Data Stream Processing Frameworks in Clouds7047.1.5.Stream Processing for Maritime Surveillance7047.1.6.Adaptive deployment for multi-cloud applications7057.1.7.Application configuration and reconfiguration in multi-cloud environments7057.1.8.Adaptive resource management for high-performance, multi-sensor systems7067.2.1.Energy Models7067.2.2.Exploiting Renewable Energy in Clouds7077.2.4.Involving Users in Energy Saving7077.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds708		6.4. PaaSage Adapter	701
6.6.SimGrid7026.7.DiFFuSE7037.New Results7037.1.Scaling Clouds7037.1.1.Fog Computing7037.1.2.Edge Cloud7037.1.3.Community Clouds7047.1.4.Evaluation of Data Stream Processing Frameworks in Clouds7047.1.5.Stream Processing for Maritime Surveillance7047.1.6.Adaptive deployment for multi-cloud applications7057.1.7.Application configuration and reconfiguration in multi-cloud environments7057.1.8.Adaptive resource management for high-performance, multi-sensor systems7067.2.1.Energy Models7067.2.2.Exploiting Renewable Energy in Clouds7067.2.3.Smart Grids7077.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds708		6.5. SAIDS	702
6.7. DifFuSE     703       7. New Results     703       7.1. Scaling Clouds     703       7.1. Scaling Clouds     703       7.1. Scaling Clouds     703       7.1.1. Fog Computing     703       7.1.2. Edge Cloud     703       7.1.3. Community Clouds     704       7.1.4. Evaluation of Data Stream Processing Frameworks in Clouds     704       7.1.5. Stream Processing for Maritime Surveillance     704       7.1.6. Adaptive deployment for multi-cloud applications     705       7.1.7. Application configuration and reconfiguration in multi-cloud environments     705       7.1.8. Adaptive resource management for high-performance, multi-sensor systems     705       7.2.1. Energy Models     706       7.2.2. Exploiting Renewable Energy in Clouds     706       7.2.3. Smart Grids     707       7.3.1. Securing Clouds     707       7.3.1. Security Monitoring in Clouds     707       7.3.2. Risk assessment in clouds     708		6.6. SIMUrid	702
7.New Results7037.1.Scaling Clouds7037.1.1.Fog Computing7037.1.2.Edge Cloud7037.1.3.Community Clouds7047.1.4.Evaluation of Data Stream Processing Frameworks in Clouds7047.1.5.Stream Processing for Maritime Surveillance7047.1.6.Adaptive deployment for multi-cloud applications7057.1.7.Application configuration and reconfiguration in multi-cloud environments7057.1.8.Adaptive resource management for high-performance, multi-sensor systems7067.2.1.Energy Models7067.2.2.Exploiting Renewable Energy in Clouds7067.2.3.Smart Grids7077.3.4.Involving Users in Energy Saving7077.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds708	-	0.7. DIFFUSE	703
7.1.Sealing Clouds7037.1.1.Fog Computing7037.1.2.Edge Cloud7037.1.3.Community Clouds7047.1.4.Evaluation of Data Stream Processing Frameworks in Clouds7047.1.5.Stream Processing for Maritime Surveillance7047.1.6.Adaptive deployment for multi-cloud applications7057.1.7.Application configuration and reconfiguration in multi-cloud environments7057.1.8.Adaptive resource management for high-performance, multi-sensor systems7067.2.1.Energy Models7067.2.2.Exploiting Renewable Energy in Clouds7067.2.3.Smart Grids7077.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds7077.3.2.Risk assessment in clouds708	7.	7.1 Sealing Clouds	703
7.1.1.Fog Computing7037.1.2.Edge Cloud7037.1.3.Community Clouds7047.1.4.Evaluation of Data Stream Processing Frameworks in Clouds7047.1.5.Stream Processing for Maritime Surveillance7047.1.6.Adaptive deployment for multi-cloud applications7057.1.7.Application configuration and reconfiguration in multi-cloud environments7057.1.8.Adaptive resource management for high-performance, multi-sensor systems7067.2.1.Energy Models7067.2.2.Exploiting Renewable Energy in Clouds7067.2.3.Smart Grids7077.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds708		7.1. Scaling Clouds	703
7.1.2.Edge Cloud7037.1.3.Community Clouds7047.1.4.Evaluation of Data Stream Processing Frameworks in Clouds7047.1.5.Stream Processing for Maritime Surveillance7047.1.6.Adaptive deployment for multi-cloud applications7057.1.7.Application configuration and reconfiguration in multi-cloud environments7057.1.8.Adaptive resource management for high-performance, multi-sensor systems7057.2.1.Energy Models7067.2.2.Exploiting Renewable Energy in Clouds7067.2.3.Smart Grids7077.3.4.Involving Users in Energy Saving7077.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds708		7.1.1. Fog Computing	703
7.1.3.Community Crouds7047.1.4.Evaluation of Data Stream Processing Frameworks in Clouds7047.1.5.Stream Processing for Maritime Surveillance7047.1.5.Stream Processing for Maritime Surveillance7047.1.6.Adaptive deployment for multi-cloud applications7057.1.7.Application configuration and reconfiguration in multi-cloud environments7057.1.8.Adaptive resource management for high-performance, multi-sensor systems7057.2.Greening Clouds7067.2.1.Energy Models7067.2.2.Exploiting Renewable Energy in Clouds7067.2.3.Smart Grids7077.3.4.Involving Users in Energy Saving7077.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds708		7.1.2. Edge Clouds	703
7.1.4.Evaluation of Data Stream Processing Frameworks in Clouds7047.1.5.Stream Processing for Maritime Surveillance7047.1.6.Adaptive deployment for multi-cloud applications7057.1.7.Application configuration and reconfiguration in multi-cloud environments7057.1.8.Adaptive resource management for high-performance, multi-sensor systems7057.2.Greening Clouds7067.2.1.Energy Models7067.2.2.Exploiting Renewable Energy in Clouds7067.2.3.Smart Grids7077.2.4.Involving Users in Energy Saving7077.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds708		7.1.5. Community Clouds 7.1.4. Evaluation of Data Stream Processing Frameworks in Clouds	704
7.1.5.Stream Processing for Warthine Surveinance7047.1.6.Adaptive deployment for multi-cloud applications7057.1.7.Application configuration and reconfiguration in multi-cloud environments7057.1.8.Adaptive resource management for high-performance, multi-sensor systems7057.2.Greening Clouds7067.2.1.Energy Models7067.2.2.Exploiting Renewable Energy in Clouds7067.2.3.Smart Grids7077.2.4.Involving Users in Energy Saving7077.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds708		7.1.4. Evaluation of Data Stream Processing Frameworks in Clouds	704
7.1.0.Adaptive deployment for hum-croud appreations7037.1.7.Application configuration and reconfiguration in multi-cloud environments7057.1.8.Adaptive resource management for high-performance, multi-sensor systems7057.2.Greening Clouds7067.2.1.Energy Models7067.2.2.Exploiting Renewable Energy in Clouds7067.2.3.Smart Grids7077.2.4.Involving Users in Energy Saving7077.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds708		7.1.5. Stream Processing for Manufile Survemance	704
7.1.7.Application comgutation and recomgutation in multi-cloud civitoninents7037.1.8.Adaptive resource management for high-performance, multi-sensor systems7057.2.Greening Clouds7067.2.1.Energy Models7067.2.2.Exploiting Renewable Energy in Clouds7067.2.3.Smart Grids7077.3.4.Involving Users in Energy Saving7077.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds708		7.1.0. Adaptive deployment for multi-cloud applications 7.1.7 Application configuration and reconfiguration in multi-cloud environments	705
7.2.Greening Clouds7067.2.1.Energy Models7067.2.2.Exploiting Renewable Energy in Clouds7067.2.3.Smart Grids7077.2.4.Involving Users in Energy Saving7077.3.Securing Clouds7077.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds708		7.1.8 Adaptive resource management for high-performance, multi-sensor systems	705
7.2.1. Energy Models7067.2.2. Exploiting Renewable Energy in Clouds7067.2.3. Smart Grids7077.2.4. Involving Users in Energy Saving7077.3. Securing Clouds7077.3.1. Security Monitoring in Clouds7077.3.2. Risk assessment in clouds708		7.2 Greening Clouds	705
7.2.2.Exploiting Renewable Energy in Clouds7067.2.3.Smart Grids7077.2.4.Involving Users in Energy Saving7077.3.Securing Clouds7077.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds708		7.2. Fnergy Models	706
7.2.3.Smart Grids7077.2.4.Involving Users in Energy Saving7077.3.Securing Clouds7077.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds708		7.2.7. Exploiting Renewable Energy in Clouds	706
7.2.4.Involving Users in Energy Saving7077.3.Securing Clouds7077.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds708		7.2.3 Smart Grids	707
7.3.Securing Clouds7077.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds708		7.2.4. Involving Users in Energy Saving	707
7.3.1.Security Monitoring in Clouds7077.3.2.Risk assessment in clouds708		7.3. Securing Clouds	707
7.3.2. Risk assessment in clouds 708		7.3.1. Security Monitoring in Clouds	707
		7.3.2. Risk assessment in clouds	708

	7.3.3. Personal Data Management in Cloud-based IoT Systems	709
	7.4. Experimenting with Clouds	709
	7.4.1. Simulation	709
	7.4.2. Use cases	710
	7.4.2.1. Simulation framework for studying between-herd pathogen spread in a region	710
	7.4.2.2. FluxNet and AmeriFlux Data Analysis	710
8.	Bilateral Contracts and Grants with Industry	. 710
	8.1. Bilateral Contracts with Industry	710
	8.2. Bilateral Grants with Industry	710
	8.2.1. Thales Research and Technology (2016-2018)	710
	8.2.2. Nokia (2015-2018)	711
9.	Partnerships and Cooperations	711
	9.1. Regional Initiatives	711
	9.1.1. CominLabs EPOC project (2013-2017)	711
	9.1.2. INDIC - Cybersecurity Pole of Excellence (2014-2020)	711
	9.2. National Initiatives	712
	9.2.1. ADEME RennesGrid	712
	9.2.2. Inria ADT SaaP (2016-2018)	712
	9.2.3. Inria ADT DiFFuSE (2017-2018)	712
	9.2.4. Inria IPL Discovery (2015-2019)	712
	9.2.5. Inria IPL CityLab (2015-2018)	713
	9.2.6. Inria IPL Hac Specis (2016-2020)	713
	9.2.7. COSMIC PRE (2016 - 2018)	713
	9.2.8. MIHMES ANR Investissements d'Avenir (2012 - 2017)	714
	9.2.9. SESAME ASTRID project (2016-2019)	714
	9.2.10. PIA ELCI (2015-2018)	714
	9.2.11. CNRS GDS EcoInfo	714
	9.3. European Initiatives	714
	9.3.1. FP7 & H2020 Projects	714
	9.3.1.1. FogGuru	714
	9.3.1.2. Fed4Fire+ (2017-2022)	715
	9.3.2. Collaborations in European Programs, Except FP7 & H2020	716
	9.3.3. Collaborations with Major European Organizations	716
	9.4. International Initiatives	717
	9.4.1. Inria International Labs	717
	9.4.2. Inria Associate Teams Not Involved in an Inria International Labs	717
	9.4.3. Inria International Partners	717
	9.5. International Research Visitors	718
	9.5.1. Visits of International Scientists	718
	9.5.2. Visits to International Teams	718
10.	Dissemination	. 719
	10.1. Promoting Scientific Activities	719
	10.1.1. Scientific Events Organisation	719
	10.1.1.1. General Chair, Scientific Chair	719
	10.1.1.2. Member of the Organizing Committees	719
	10.1.2. Scientific Events Selection	719
	10.1.2.1. Chair of Conference Program Committees	719
	10.1.2.2. Member of the Conference Program Committees	719
	10.1.3. Journal	719
	10.1.3.1. Member of the Editorial Boards	719
	10.1.3.2. Reviewer - Reviewing Activities	720

10.1.4. Invited Talks	720
10.1.5. Leadership within the Scientific Community	720
10.1.6. Scientific Expertise	720
10.1.7. Research Administration	721
10.2. Teaching - Supervision - Juries	721
10.2.1. Teaching	721
10.2.2. Supervision	722
10.2.3. Juries	723
10.3. Popularization	724
11. Bibliography	

# **Project-Team MYRIADS**

*Creation of the Team: 2010 January 01, updated into Project-Team: 2012 January 01* **Keywords:** 

# **Computer Science and Digital Science:**

- A1.1.4. High performance computing
- A1.1.5. Exascale
- A1.1.6. Cloud
- A1.1.7. Peer to peer
- A1.1.9. Fault tolerant systems
- A1.1.13. Virtualization
- A1.2. Networks
- A1.2.4. QoS, performance evaluation
- A1.2.5. Internet of things
- A1.3. Distributed Systems
- A1.6. Green Computing
- A2.1.7. Distributed programming
- A2.2.3. Run-time systems
- A2.3.2. Cyber-physical systems
- A2.4.2. Model-checking
- A2.6. Infrastructure software
- A2.6.1. Operating systems
- A2.6.2. Middleware
- A2.6.3. Virtual machines
- A3.1.2. Data management, quering and storage
- A3.1.3. Distributed data
- A4.7. Access control
- A4.9. Security supervision
- A4.9.1. Intrusion detection
- A4.9.3. Reaction to attacks
- A5.6. Virtual reality, augmented reality
- A6.1.3. Discrete Modeling (multi-agent, people centered)
- A6.2.6. Optimization
- A6.2.7. High performance computing
- A7.1. Algorithms
- A8.2. Optimization

# **Other Research Topics and Application Domains:**

- B2.3. Epidemiology
- B3.1. Sustainable development
- B3.2. Climate and meteorology
- B4.3. Renewable energy production
- B4.4. Energy delivery

B4.4.1. - Smart grids B4.5. - Energy consumption

- B4.5.1. Green computing
- B5.1. Factory of the future
- B5.8. Learning and training
- B6.1. Software industry
- B6.1.1. Software engineering
- B6.3. Network functions
- B6.3.3. Network Management
- B6.4. Internet of things
- B6.5. Information systems
- B6.6. Embedded systems
- B8.1. Smart building/home
- B8.2. Connected city
- B8.5. Smart society
- B9.1. Education
- B9.1.1. E-learning, MOOC
- B9.1.2. Serious games
- B9.4.1. Computer science
- B9.6. Reproducibility
- B9.7. Knowledge dissemination
- B9.7.1. Open access
- B9.7.2. Open data
- B9.8. Privacy

# 1. Personnel

#### **Research Scientists**

Christine Morin [Team leader, Inria, Senior Researcher, HDR] Deborah Agarwal [Inria, Senior Researcher, Advanced Research Position] Jean-Pierre Banâtre [Univ de Rennes I, Emeritus, until Oct 2017] Anne-Cécile Orgerie [CNRS, Researcher]

#### **Faculty Members**

Nikos Parlavantzas [INSA Rennes, Associate Professor] Jean-Louis Pazat [INSA Rennes, Professor, HDR] Guillaume Pierre [Univ de Rennes I, Professor] Martin Quinson [Ecole normale supérieure de Rennes, Professor, HDR] Cédric Tedeschi [Univ de Rennes I, Associate Professor, HDR]

# **External Collaborator**

Louis Rilling [DGA]

#### **Technical Staff**

Subarna Chatterjee [Inria, Feb 2017] Toufik Boubehziz [Inria, from Mar 2017] Yue Li [Inria, from Mar 2017] Manh Linh Pham [Inria] Sébastien Philippot [Inria] Subhadeep Sarkar [Inria, from Feb 2017] Alexandre Van Kempen [Inria, until Feb 2017] Pascal Morillon [Univ. Rennes I, Engineer] Matthieu Simonin [Inria, Research engineer] David Margery [Inria, Research engineer, part-time]

#### **PhD Students**

Arif Ahmed [Univ de Rennes I] Mehdi Belkhiria [Univ de Rennes I, from Dec 2017] Yasmina Bouizem [Tlemcen University, from Nov 2017] Clément El Baz [Inria, from Oct 2017] Ali Fahs [Univ de Rennes I, from Oct 2017] Anna Giannakou [Inria, until May 2017] Baptiste Goupille–Lescar [Thales] Loic Guegan [Ecole normale supérieure de Rennes, from Oct 2017] Baptiste Goupille-Lescar [Thales, PhD Student, granted by CIFRE] David Guyon [Univ de Rennes I] Sabbir Hasan Rochi [INSA Rennes, PhD Student, until May 2017] Yunbo Li [Inria, until Feb 2017] The Anh Pham [Inria] Genc Tato [Inria] Amir Teshome Wonjiga [Inria]

**Post-Doctoral Fellows** 

Ehsan Ahvar [Inria, from Jun 2017] Benjamin Camus [Inria] Subarna Chatterjee [Inria, from Mar 2017]

Administrative Assistant Maryse Fouche [Inria]

# 2. Overall Objectives

# 2.1. General Objectives

MYRIADS is a joint team with INRIA, CNRS, UNIVERSITY RENNES 1, INSA RENNES and ENS RENNES. It is part of IRISA (D1 department on large scale systems) and INRIA RENNES – BRETAGNE ATLANTIQUE.

The objective of MYRIADS is to design and implement systems for autonomous service and resource management in interconnected and distributed clouds. The team tackles the challenges of dependable application execution and efficient resource management in highly distributed clouds.

# 2.2. Context

The MYRIADS team research activities are conducted in the context of the future of Internet.

Internet of Services. Myriads of applications are provided to more than one billion users <sup>0</sup> all over the world. Over time, these applications are becoming more and more sophisticated, a given application being a composition of services likely to be executed on various sites located in different geographical locations. The Internet of Services is spreading all domains: home, administration, business, industry and science. Everyone is involved in the Internet of Services: citizens, enterprises, scientists are application, service and resource consumers and/or providers over the Internet.

<sup>&</sup>lt;sup>0</sup>According to World Stats, there are 3.67 billion Internet users i.e. more than half of the total world population in June 2016 http:// www.internetworldstats.com/stats.htm.

- Outsourcing. Software is provided as a service over the Internet. Myriads of applications are available on-line to billions of users as, for instance, *GoogleApps* (Gmail). After decades in which companies used to host their entire IT infrastructures in-house, a major shift is occurring where these infrastructures are outsourced to external operators such as Data Centers and Computing Clouds. In the Internet of Services, not only software but also infrastructure are delivered as a service. Clouds turned computing and storage into a utility. Just like water or electricity, they are available in virtually infinite amounts and their consumption can be adapted within seconds like opening or closing a water tap. The main transition, however, is the change in business models. Companies or scientists do not need to buy and operate their own data centers anymore. Instead, the compute and storage hardware investments before starting a business. Even more, the new model allows users to adapt their resources within minutes, e.g., scale up to handle peak loads or rent large numbers of computers for a short experiment. The risk of wasting money by either under-utilization or undersized data centers is shifted from the user to the provider.
- Sharing and Cooperation. Sharing information and cooperating over the Internet are also important user needs both in the private and the professional spheres. This is exemplified by various services that have been developed in the last decade. Peer-to-peer networks are extensively used by citizens in order to share musics and movies. A service like *Flickr* allowing individuals to share pictures is also very popular. Social networks such as *FaceBook* or *LinkedIn* link millions of users who share various kinds of information within communities. Virtual organizations tightly connected to Grids allow scientists to share computing resources aggregated from different institutions (universities, computing centers...). The EGEE European Grid is an example of production Grid shared by thousands of scientists all over Europe.

# 2.3. Challenges

The term cloud was coined 10 years ago. Today cloud computing is widely adopted for a wide range of usage: information systems outsourcing, web service hosting, scientific computing, data analytics, back-end of mobile and IoT applications. There is a wide variety of cloud service providers (IaaS, PaaS, SaaS) resulting in difficulties for customers to select the services fitting their needs. Production clouds are powered by huge data centers that customers reach through the Internet. This current model raises a number of issues. Cloud computing generates a lot of traffic resulting in ISP providers needing to increase the network capacity. An increasing amount of always larger data centers consumes a lot of energy. Cloud customers experience poor quality of experience for highly interactive mobile applications as their requests are dealt with in data centers that are several hops away. The centralization of data in clouds also raises (i) security issues as clouds are a target of choice for attackers and (ii) privacy issues with data aggregation. Recently new cloud architectures have been proposed to overcome the scalability, latency, and energy issues of traditional centralized data centers. Various flavors of distributed cloud computing are emerging depending on the resources exploited: resources in the core network (distributed cloud), resources at the edge of the network (edge clouds) and even resources in the people swarms of devices (fog computing) enabling scalable cloud computing. These distributed clouds raise new challenges for resource and application management.

The ultimate goal of Myriads team is making highly distributed clouds sustainable. By sustainability we mean green, efficient and secure clouds. We plan to study highly distributed clouds including edge clouds and fog computing. In this context, we will investigate novel techniques for greening clouds including the optimization of energy consumption in distributed clouds in the context of smart grids. As more and more critical information system are outsourced in the cloud and personal data captured by sensors embedded in smart objects and smartphones are stored in the cloud, we will investigate security and privacy issues in two directions: cloud security monitoring and personal data protection in cloud-based IoT applications.

System research requires experimental validation based on simulation and/or prototyping. Reproducible experimentation is essential. We will contribute to the design and implementation of simulators well suited to the study of distributed clouds (architecture, energy consumption) and of large scale experimentation platforms for distributed systems enabling reproducible experiments.

# 3. Research Program

# 3.1. Introduction

In this section, we present our research challenges along four work directions: resource and application management in distributed cloud architectures for scaling clouds in Section 3.2, energy management strategies for greening clouds in Section 3.3, security and data protection aspects for securing cloud-based information systems and applications in Section 3.4, and methods for experimenting with clouds in Section 3.5.

# 3.2. Scaling clouds

#### 3.2.1. Resource management in hierarchical clouds

The next generation of utility computing appears to be an evolution from highly centralized clouds towards more decentralized platforms. Today, cloud computing platforms mostly rely on large data centers servicing a multitude of clients from the edge of the Internet. Servicing cloud clients in this manner suggests that locality patterns are ignored: wherever the client issues his/her request from, the request will have to go through the backbone of the Internet provider to the other side of the network where the data center relies. Besides this extra network traffic and this latency overhead that could be avoided, other common centralization drawbacks in this context stand in limitations in terms of security/legal issues and resilience.

At the same time, it appears that network backbones are over-provisioned for most of their usage. This advocates for placing computing resources directly within the backbone network. The general challenge of resource management for such clouds stands in trying to be locality-aware: for the needs of an application, several virtual machines may exchange data. Placing them *close* to each others can significantly improve the performance of the application they compose. More generally, building an overlay network which takes the hierarchical aspects of the platform without being a hierarchical overlay – which comes with load balancing and resilience issues is a challenge by itself.

The results of these works are planned to be integrated into the Discovery initiative [60] which aims at revisiting OpenStack to offer a cloud stack able to manage utility computing platforms where computing resources are located in small computing centers in the backbone's PoPs (Point of Presence) and interconnected through the backbone's internal links.

#### 3.2.2. Resource management in mobile edge clouds

Mobile edge cloud (MEC) infrastructures are composed of compute, storage and networking resources located at the edge of wide-area networks, in immediate proximity to the end users. Instead of treating the mobile operator's network as a high-latency dumb pipe between the end users and the external service providers, MEC platforms aim at deploying cloud functionalities *within* the mobile phone network, inside or close to the mobile access points. Doing so is expected to deliver added value to the content providers and the end users by enabling new types of applications ranging from Internet-of-Things applications to extremely interactive systems (e.g., augmented reality). Simultaneously, it will generate extra revenue streams for the mobile network operators, by allowing them to position themselves as cloud computing operators and to rent their already-deployed infrastructure to content and application providers.

Mobile edge clouds have very different geographical distribution compared to traditional clouds. While traditional clouds are composed of many reliable and powerful machines located in a very small number of data centers and interconnected by very high-speed networks, mobile edge cloud are composed of a very large number of points-of-presence with a couple of weak and potentially unreliable servers, interconnected with each other by commodity long-distance networks. This creates new demands for the organization of a scalable mobile edge computing infrastructure, and opens new directions for research.

The main challenges that we plan to address are:

- How should an edge cloud infrastructure be designed such that it remains scalable, fault-tolerant, controllable, energy-efficient, etc.?
- How should applications making use of edge clouds be organized? One promising direction is to explore the extent to which stream-data processing platforms such as Apache Spark and Apache Flink can be adapted to become one of the main application programming paradigms in such environments.

## 3.2.3. Self-optimizing applications in multi-cloud environments

As the use of cloud computing becomes pervasive, the ability to deploy an application on a multi-cloud infrastructure becomes increasingly important. Potential benefits include avoiding dependence on a single vendor, taking advantage of lower resource prices or resource proximity, and enhancing application availability. Supporting multi-cloud application management involves two tasks. First, it involves selecting an initial multicloud application deployment that best satisfies application objectives and optimizes performance and cost. Second, it involves dynamically adapting the application deployment in order to react to changes in execution conditions, application objectives, cloud provider offerings, or resource prices. Handling price changes in particular is becoming increasingly complex. The reason is the growing trend of providers offering sophisticated, dynamic pricing models that allow buying and selling resources of finer granularities for shorter time durations with varying prices.

Although multi-cloud platforms are starting to emerge, these platforms impose a considerable amount of effort on developers and operations engineers, provide no support for dynamic pricing, and lack the responsiveness and scalability necessary for handling highly-distributed, dynamic applications with strict quality requirements. The goal of this work is to develop techniques and mechanisms for automating application management, enabling applications to cope with and take advantage of the dynamic, diverse, multi-cloud environment in which they operate.

The main challenges arising in this context are:

- selecting effective decision-making approaches for application adaptation,
- supporting scalable monitoring and adaptation across multiple clouds,
- performing adaptation actions in a cost-efficient and safe manner.

# **3.3.** Greening clouds

ICT (Information and Communications Technologies) ecosystem now approaches 5% of world electricity consumption and this ICT energy use will continue grow fast because of the information appetite of Big Data, big networks and big infrastructures as Clouds that unavoidably leads to big power.

## 3.3.1. Smart grids and clouds

We propose exploiting Smart Grid technologies to come to the rescue of energy-hungry Clouds. Unlike in traditional electrical distribution networks, where power can only be moved and scheduled in very limited ways, Smart Grids dynamically and effectively adapt supply to demand and limit electricity losses (currently 10% of produced energy is lost during transmission and distribution).

For instance, when a user submits a Cloud request (such as a Google search for instance), it is routed to a data center that processes it, computes the answer and sends it back to the user. Google owns several data centers spread across the world and for performance reasons, the center answering the user's request is more likely to be the one closest to the user. However, this data center may be less energy efficient. This request may have consumed less energy, or a different kind of energy (renewable or not), if it had been sent to this further data center. In this case, the response time would have been increased but maybe not noticeably: a different trade-off between quality of service (QoS) and energy-efficiency could have been adopted.

While Clouds come naturally to the rescue of Smart Grids for dealing with this big data issue, little attention has been paid to the benefits that Smart Grids could bring to distributed Clouds. To our knowledge, no previous work has exploited the Smart Grids potential to obtain and control the energy consumption of entire Cloud infrastructures from underlying facilities such as air conditioning equipment (which accounts for 30% to 50% of a data center's electricity bill) to network resources (which are often operated by several actors) and to computing resources (with their heterogeneity and distribution across multiple data centers). We aim at taking advantage of the opportunity brought by the Smart Grids to exploit renewable energy availability and to optimize energy management in distributed Clouds.

#### 3.3.2. Energy cost models

Cloud computing allows users to outsource the computer resources required for their applications instead of using a local installation. It offers on-demand access to the resources through the Internet with a pay-as-you-go pricing model. However, this model hides the electricity cost of running these infrastructures.

The costs of current data centers are mostly driven by their energy consumption (specifically by the air conditioning, computing and networking infrastructure). Yet, current pricing models are usually static and rarely consider the facilities' energy consumption per user. The challenge is to provide a fair and predictable model to attribute the overall energy costs per virtual machine and to increase energy-awareness of users.

Another goal consists in better understanding the energy consumption of computing and networking resources of Clouds in order to provide energy cost models for the entire infrastructure including incentivizing cost models for both Cloud providers and energy suppliers. These models will be based on experimental measurement campaigns on heterogeneous devices. Inferring a cost model from energy measurements is an arduous task since simple models are not convincing, as shown in our previous work. We aim at proposing and validating energy cost models for the heterogeneous Cloud infrastructures in one hand, and the energy distribution grid on the other hand. These models will be integrated into simulation frameworks in order to validate our energy-efficient algorithms at larger scale.

#### 3.3.3. Energy-aware users

In a Cloud moderately loaded, some servers may be turned off when not used for energy saving purpose. Cloud providers can apply resource management strategies to favor idle servers. Some of the existing solutions propose mechanisms to optimize VM scheduling in the Cloud. A common solution is to consolidate the mapping of the VMs in the Cloud by grouping them in a fewer number of servers. The unused servers can then be turned off in order to lower the global electricity consumption.

Indeed, current work focuses on possible levers at the virtual machine suppliers and/or services. However, users are not involved in the choice of using these levers while significant energy savings could be achieved with their help. For example, they might agree to delay slightly the calculation of the response to their applications on the Cloud or accept that it is supported by a remote data center, to save energy or wait for the availability of renewable energy. The VMs are black boxes from the Cloud provider point of view. So, the user is the only one to know the applications running on her VMs.

We plan to explore possible collaborations between virtual machine suppliers, service providers and users of Clouds in order to provide users with ways of participating in the reduction of the Clouds energy consumption. This work will follow two directions: 1) to investigate compromises between power and performance/service quality that cloud providers can offer to their users and to propose them a variety of options adapted to their workload; and 2) to develop mechanisms for each layer of the Cloud software stack to provide users with a quantification of the energy consumed by each of their options as an incentive to become greener.

# 3.4. Securing clouds

# 3.4.1. Security monitoring SLO

While the trend for companies to outsource their information system in clouds is confirmed, the problem of securing an information system becomes more difficult. Indeed, in the case of infrastructure clouds, physical

resources are shared between companies (also called tenants) but each tenant controls only parts of the shared resources, and, thanks to virtualization, the information system can be dynamically and automatically reconfigured with added or removed resources (for example starting or stopping virtual machines), or even moved between physical resources (for example using virtual machine migration). Partial control of shared resources brings new classes of attacks between tenants, and security monitoring mechanisms to detect such attacks are better placed out of the tenant-controlled virtual information systems, that is under control of the cloud provider. Dynamic and automatic reconfigurations of the information system make it unfeasible for a tenant's security administrator to setup the security monitoring components to detect attacks, and thus an automated self-adaptable security monitoring service is required.

Combining the two previous statements, there is a need for a dependable, automatic security monitoring service provided to tenants by the cloud provider. Our goal is to address the following challenges to design such a security monitoring service:

- 1. to define relevant Service-Level Objectives (SLOs) of a security monitoring service, that can figure in the Service-Level Agreement (SLA) signed between a cloud provider and a tenant;
- 2. to design heuristics to automatically configure provider-controlled security monitoring software components and devices so that SLOs are reached, even during automatic reconfigurations of tenants' information systems;
- 3. to design evaluation methods for tenants to check that SLOs are reached.

Moreover in challenges 2 and 3 the following sub-challenges must be addressed:

- although SLAs are bi-lateral contracts between the provider and each tenant, the implementation of the contracts is based on shared resources, and thus we must study methods to combine the SLOs;
- the designed methods should have a minimal impact on performance.

# 3.4.2. Data Protection in Cloud-based IoT Services

The Internet of Things is becoming a reality. Individuals have their own swarm of connected devices (e.g. smartphone, wearables, and home connected objects) continually collecting personal data. A novel generation of services is emerging exploiting data streams produced by the devices' sensors. People are deprived of control of their personal data as they don't know precisely what data are collected by service providers operating on Internet (oISP), for which purpose they could be used, for how long they are stored, and to whom they are disclosed. In response to privacy concerns the European Union has introduced, with the Global Data Protection Regulation (GDPR), new rules aimed at enforcing the people's rights to personal data protection. The GDPR also gives strong incentives to oISPs to comply. However, today, oISPs can't make their systems GDPR-compliant since they don't have the required technologies. We argue that a new generation of system is mandatory for enabling oISPs to conform to the GDPR. We plan to to design an open source distributed operating system for native implementation of new GDPR rules and ease the programming of compliant cloud-based IoT services. Among the new rules, transparency, right of erasure, and accountability are the most challenging ones to be implemented in IoT environments but could fundamentally increase people's confidence in oISPs. Deployed on individuals' swarms of devices and oISPs' cloud-hosted servers, it will enforce detailed data protection agreements and accountability of oISPs' data processing activities. Ultimately we will show to what extend the new GDPR rules can be implemented for cloud-based IoT services.

# **3.5. Experimenting with Clouds**

Cloud platforms are challenging to evaluate and study with a sound scientific methodology. As with any distributed platform, it is very difficult to gather a global and precise view of the system state. Experiments are not reproducible by default since these systems are shared between several stakeholder. This is even worsen by the fact that microscopic differences in the experimental conditions can lead to drastic changes since typical Cloud applications continuously adapt their behavior to the system conditions.

## 3.5.1. Experimentation methodologies for clouds

We propose to combine two complementary experimental approaches: direct execution on testbeds such as Grid'5000, that are eminently believable but rather labor intensive, and simulations (using *e.g.* SimGrid) that are much more light-weighted, but requires are careful assessment. One specificity of the Myriads team is that we are working on these experimental methodologies *per se*, raising the standards of *good experiments* in our community.

We plan to make SimGrid widely usable beyond research laboratories, in order to evaluate industrial systems and to teach the future generations of cloud practitioners. This requires to frame the specific concepts of Cloud systems and platforms in actionable interfaces. The challenge is to make the framework both easy to use for simple studies in educational settings while modular and extensible to suit the specific needs of every advanced industrial-class users.

We aim at leveraging the convergence opportunities between methodologies by further bridging simulation and real testbeds. The predictions obtained from the simulator should be validated against some real-world experiments obtained on the target production platform, or on a similar platform. This (in)validation of the predicted results often improves the understanding of the modeled system. On the other side, it may even happen that the measured discrepancies are due to some mis-configuration of the real platform that would have been undetected without this (in)validation study. In that sense, the simulator constitutes a precious tool for the quality assurance of real testbeds such as Grid'5000.

Scientists need more help to make there Cloud experiments fully reproducible, in the sprit of Open Science exemplified by the HAL Open Archive, actively backed by Inria. Users still need practical solutions to archive, share and compare the whole experimental settings, including the raw data production (particularly in the case of real testbeds) and their statistical analysis. This is a long lasting task to which we plan to collaborate through the research communities gathered around the Grid'5000 and SimGrid scientific instruments.

Finally, since correction and performance can constitute contradictory goals, it is particularly important to study them jointly. To that extend, we want to bridge the performance studies, that constitute our main scientific heritage, to correction studies leveraging formal techniques. SimGrid already includes to exhaustively explore the possible executions. We plan to continue this work to ease the use of the relevant formal methods to the experimenter studying Cloud systems.

# 3.5.2. Use cases

In system research it is important to work on real-world use cases from which we extract requirements inspiring new research directions and with which we can validate the system services and mechanisms we propose. In the framework of our close collaboration with the Data Science Technology department of the LBNL, we will investigate cloud usage for scientific data management. Next-generation scientific discoveries are at the boundaries of datasets, e.g., across multiple science disciplines, institutions and spatial and temporal scales. Today, data integration processes and methods are largely adhoc or manual. A generalized resource infrastructure that integrates knowledge of the data and the processing tasks being performed by the user in the context of the data and resource lifecycle is needed. Clouds provide an important infrastructure platform that can be leveraged by including knowledge for distributed data integration.

# 4. Application Domains

# **4.1.** Main Application Domains

The Myriads team investigates the design and implementation of system services. Thus its research activities address a broad range of application domains. We validate our research results with selected use cases in the following application domains:

• Smart city services,

- Smart grids,
- Energy and sustainable development,
- Home IoT applications,
- Bio-informatics applications,
- Data science applications,
- Computational science applications,
- Numerical simulations.

# 5. Highlights of the Year

# 5.1. Highlights of the Year

- The FogGuru project was accepted and started on September 1st 2017. FogGuru is a European H2020 Maria-Skłodowska-Curie Action (MSCA) European Industrial Doctorate (EID) training project which aims to to train eight talented PhD students with an innovative and intersectoral research program to constitute the next generation of European Cloud and Fog computing experts. It is coordinated by Guillaume Pierre.
- Cédric Tedeschi defended his *habilitation à diriger des recherches* summarizing his research activity of the last seven years.

# 5.1.1. Awards

- Best paper award for Timothée Haudebourg and Anne-Cécile Orgerie at the International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP 2017) for the paper entitled "On the Energy Efficiency of Sleeping and Rate Adaptation for Network Devices"
- Christine Morin has been selected to be included in the 2017 list of "N2Women:Stars in Computer Networking and Communications". The "N2Women:Stars in Computer Networking and Communications" is an annual list focusing in amazing women who have had a major impact in networking and/or communications.

BEST PAPERS AWARDS :

[34] ICA3PP 2017 - 17th International Conference on Algorithms and Architectures for Parallel Processing. T. HAUDEBOURG, A.-C. ORGERIE.

# 6. New Software and Platforms

# 6.1. ConPaaS

KEYWORDS: Cloud computing - PaaS SCIENTIFIC DESCRIPTION: Contact:

Guillaume Pierre, Guillaume.Pierre@irisa.fr URL:

http://www.conpaas.eu/ Status:

Version 1.4.2 License:

**BSD** Presentation:

ConPaaS [60] is a runtime environment for hosting applications in the cloud. It aims at offering the full power of the cloud to application developers while shielding them from the associated complexity of the cloud. ConPaaS is designed to host both high-performance scientific applications and online Web applications. It automates the entire life-cycle of an application, including collaborative development, deployment, performance monitoring, and automatic scaling. This allows developers to focus their attention on application-specific concerns rather than on cloud-specific details. Active contributors (from the Myriads team):

Eliya Buyukkaya, Ancuta Iordache, Morteza Neishaboori, Guillaume Pierre, Dzenan Softic, Genc Tato, Teodor Crivat. Impact:

ConPaaS is recognized as one of the major open-source PaaS environments. It is being developed by teams in Rennes, Amsterdan, Berlin and Ljubljana. Technology transfer of ConPaaS technology is ongoing in the context of the MC-DATA EIT ICT Labs project.

FUNCTIONAL DESCRIPTION: ConPaaS is a runtime environment for hosting applications in the cloud. It aims at offering the full power of the cloud to application developers while shielding them from the associated complexity of the cloud. ConPaaS is designed to host both high-performance scientific applications and online Web applications. It automates the entire life-cycle of an application, including collaborative development, deployment, performance monitoring, and automatic scaling. This allows developers to focus their attention on application-specific concerns rather than on cloud-specific details.

- Participants: Ancuta Iordache, Dzenan Softic, Eliya Buyukkaya, Genc Tato, Guillaume Pierre, Morteza Neishaboori and Teodor Crivat
- Contact: Guillaume Pierre
- URL: http://www.conpaas.eu/

# 6.2. GinFlow

KEYWORDS: Dynamic adaptation - Distributed Applications - Distributed - Distributed computing - Workflow - Framework

FUNCTIONAL DESCRIPTION: GinFlow decentralizes the coordination of the execution of workflow-based applications. GinFlow relyies on an architecture where multiple service agents (SA) coordinate each others through a shared space containing the workflow description and current status. GinFlow allows the user to define several variants of a workflow and to switch from one to the other during run time.

- Participants: Cédric Tedeschi, Hector Fernandez, Javier Rojas Balderrama, Matthieu Simonin and Thierry Priol
- Partner: Université de Rennes 1
- Contact: Cédric Tedeschi
- URL: http://ginflow.inria.fr

# 6.3. Merkat

FUNCTIONAL DESCRIPTION: Merkat is a platform that allows users of an organization to automatically manage and scale their applications while maximizing the infrastructure's utilization. Merkat is generic and extensible, allowing users to automate the application deployment and management process. Users have the flexibility to control how many resources are allocated to their applications and to define their own resource demand adaptation policies. Merkat applies an unique approach to multiplex the infrastructure capacity between the applications, by implementing a proportional-share market and allowing applications to adapt autonomously to resource price and their given performance objectives. The price of the acquired resources acts as a control mechanism to ensure that resources are distributed to applications according to the user's value for them. Merkat was evaluated on Grid'5000 with several scientific applications.

- Participants: Christine Morin, Nikolaos Parlavantzas and Stefania Costache
- Contact: Nikolaos Parlavantzas
- URL: http://www.irisa.fr/myriads/software/Merkat/

# 6.4. PaaSage Adapter

KEYWORDS: Cloud computing - Dynamic adaptation - Cloud applications management

FUNCTIONAL DESCRIPTION: The purpose of the Adapter is to transform the current configuration of a cloud application into a target configuration in an efficient and safe way. The Adapter is part of PaaSage, an open-source platform for modeling, deploying and executing applications on different clouds in an optimal manner. The Adapter has the following responsibilities: (1) validating reconfiguration plans, (2) applying the plans to the running system, and (3) maintaining an up-to-date representation of the current system state.

- Contact: Nikolaos Parlavantzas
- URL: https://team.inria.fr/myriads/software-and-platforms/paasage-adapter/

# 6.5. SAIDS

#### self-adaptable intrusion detection system

#### KEYWORDS: Cloud - Security

FUNCTIONAL DESCRIPTION: SAIDS is a self-adaptable intrusion detection system for IaaS clouds. To maintain an effective level of intrusion detection, SAIDS monitors changes in the virtual infrastructure of a Cloud environment and reconfigures its components (security probes) accordingly. SAIDS can also reconfigure probes in the case of a change in the list of running services.

- Authors: Anna Giannakou and Jean-Léon Cusinato
- Contact: Christine Morin

# 6.6. SimGrid

KEYWORDS: Large-scale Emulators - Grid Computing - Distributed Applications

SCIENTIFIC DESCRIPTION: SimGrid is a toolkit that provides core functionalities for the simulation of distributed applications in heterogeneous distributed environments. The simulation engine uses algorithmic and implementation techniques toward the fast simulation of large systems on a single machine. The models are theoretically grounded and experimentally validated. The results are reproducible, enabling better scientific practices.

Its models of networks, cpus and disks are adapted to (Data)Grids, P2P, Clouds, Clusters and HPC, allowing multi-domain studies. It can be used either to simulate algorithms and prototypes of applications, or to emulate real MPI applications through the virtualization of their communication, or to formally assess algorithms and applications that can run in the framework.

The formal verification module explores all possible message interleavings in the application, searching for states violating the provided properties. We recently added the ability to assess liveness properties over arbitrary and legacy codes, thanks to a system-level introspection tool that provides a finely detailed view of the running application to the model checker. This can for example be leveraged to verify both safety or liveness properties, on arbitrary MPI code written in C/C++/Fortran.

RELEASE FUNCTIONAL DESCRIPTION:

- Four releases in 2017. Major changes:
  - S4U: many progress, toward SimGrid v4.0. About 80% of the features offered by SimDag and MSG are now integrated, along with examples. Users can now write plugins to extend SimGrid.
  - SMPI: Support MPI 2.2, RMA support, Convert internals to C++.
  - Java: Massive memleaks and performance issues fixed.
  - New models: Multi-core VMs, Energy consumption due to the network
  - All internals are now converted to C++, and most of our internally developped data containers were replaced with std::\* constructs.
  - (+ bug fixes, cleanups and documentation improvements)
- Participants: Adrien Lèbre, Arnaud Legrand, Augustin Degomme, Florence Perronnin, Frédéric Suter, Jean-Marc Vincent, Jonathan Pastor, Jonathan Rouzaud-Cornabas, Luka Stanisic, Mario Südholt and Martin Quinson
- Partners: CNRS ENS Rennes
- Contact: Martin Quinson
- URL: http://simgrid.gforge.inria.fr/

# 6.7. DiFFuSE

Distributed framework for cloud-based epidemic simulations KEYWORDS: Simulation - Cloud

FUNCTIONAL DESCRIPTION: The DiFFuSE framework enables simulations of epidemics to take full advantage of cloud environments. The framework provides design support, reusable code, and tools for building and executing epidemic simulations. Notably, the framework automatically handles failures and supports elastic allocation of resources from multiple clouds.

- Contact: Nikolaos Parlavantzas
- URL: https://team.inria.fr/myriads/software-and-platforms/diffuse/

# 7. New Results

# 7.1. Scaling Clouds

#### 7.1.1. Fog Computing

**Participants:** Guillaume Pierre, Arif Ahmed, Ali Fahs, Alexandre Van Kempen, Salsabil Amri, Vinothkumar Nagasayanan, Berenger Nguyen Nhon.

Fog computing aims to extend datacenter-based cloud platforms with additional compute, networking and storage resources located in the immediate vicinity of the end users. By bringing computation where the input data was produced and the resulting output data will be consumed, fog computing is expected to support new types of applications which either require very low network latency (e.g., augmented reality applications) or which produce large data volumes which are relevant only locally (e.g., IoT-based data analytics).

Fog computing architectures are fundamentally different from those of traditional cloud platforms: to provide computing resources in physical proximity of any end user, fog computing platforms must necessarily rely on very large numbers of small Points-of-Presence connected to each other with commodity networks whereas clouds are typically organized with a handful of extremely powerful data centers connected by dedicated ultrahigh-speed networks. This geographical spread also implies that the machines used in any Point-of-Presence may not be datacenter-grade servers but much weaker commodity machines.

We investigated the challenges of efficiently deploying Docker containers in fog platforms composed of tiny single-board computers such as Raspberry PIs. This operation can be painfully slow, in the order of multiple minutes depending on the container's image size and network condition. We showed that this bad performance is not only due to hardware limitations, but it is largely due to inefficiencies in the way Docker implements the container's image download operation. We proposed a number of optimization techniques which , when combined together, make container deployment up to 4 times faster than the vanilla Docker implementation. A publication on this topic is under submission.

Although fog computing infrastructures are fundamentally distributed, their management part still remains centralized: a single node (or small group of nodes) is in charge of maintaining the list of available server machines, monitoring them, distributing software to them, deciding which server must take care of which task, etc. We therefore aim to reduce the discrepancy between the broadly distributed compute/storage resources and the – currently – extremely centralized control of these resources, by focusing first on the resource scheduling function. This project has just started, and we expect to obtain the first results in 2018.

# 7.1.2. Edge Cloud

Participants: Anne-Cécile Orgerie, Cédric Tedeschi, Matthieu Simonin, Ehsan Ahvar, Genc Tato.

Myriads is involved in the Discovery project, whose goal is to design, develop and experiment a software stack for a distributed cloud platform where resources are directly injected into the backbone of the network [60]. To this end, we designed a novel family of overlay network to operate messaging and routing on top of such a distributed utility computing platform. The big picture of these overlays was described in a workshop [47].

## 7.1.3. Community Clouds

#### Participant: Jean-Louis Pazat.

In this work we consider an infrastructure based on devices (such as Internet boxes and NAS) owned and operated by end-users. A typical use-case is the sharing of CPU and storage capabilities by a community of users. This sharing is operated by hosting services to local and remote users. The devices of this distributed infrastructure have heterogeneous capabilities and no guaranteed availability. It is therefore challenging to ensure to the guest service a minimal hosting service level, such as availability or QoS.

We consider services build as an application based on micro-services. Such an application is deployed on the infrastructure by instantiating its constituant micro-services on some devices. One micro-services may rely on others micro-services to enable its own service. The performance of the resulting application is therefore highly dependent from the placement for each micro-service instance. Device parameters like CPU capabilities or network bandwidth and latency have a significant impact on the resulting response time of the micro-service, hence the application.

We explore solutions to adapt the placement of the micro-services to the capabilities of the infrastructure. As a first step, we are studying a static system where these capabilities are not variating. The placement decision can be expressed as the solution of an NP-Complete optimization problem. We have shown that a solution for this problem can be found with reasonably good precision using a meta-heuristic called Particle Swarm Optimization. The next step will be to study how this solution can be adapted in a dynamic system by considering the variations of the CPU and Network parameters and the availability of the devices.

This work in done in the context of Bruno Stevant's PhD thesis co-advised by Jean-Louis Pazat (Bruno Stevant is a member of OCIF team).

#### 7.1.4. Evaluation of Data Stream Processing Frameworks in Clouds

Participants: Christine Morin, Deborah Agarwal, Subarna Chatterjee.

We address the problem of selecting a correct stream processing framework for a given application to be executed within a specific physical infrastructure. For this purpose, we have performed a thorough comparative analysis of three data stream processing platforms - Apache Flink, Apache Storm, and Twitter Heron (the enhanced version of Apache Storm), that are chosen based on their potential to process both streams and batches in real-time. For the comparative performance analysis of the chosen platforms, we have experimented using 8-node clusters on Grid5000 experimentation testbed and have selected a wide variety of applications ranging from a conventional benchmark (word count application) to sensor-based IoT application (air quality monitoring application) and statistical batch processing application (flight delay analysis application). The work focuses to analyze the performance of the frameworks in terms of the volume and throughput of data streams that each framework can possibly handle. The impact of each framework on the operating system is analyzed by experimenting and studying the resource utilization of the platforms in terms of CPU utilization, memory consumption. The energy consumption of the platforms is also studied to understand the suitability of the platforms towards green computing. Last, but not the least, the fault tolerance of the frameworks is also studied and analyzed. Lessons learnt from this work will precisely enlighten IaaS cloud end-users to wisely choose the correct streaming platform in order to run a particular application within a given set of VMs and will assist the cloud-providers to rationally allocate VMs equipped with a particular stream processing framework to PaaS cloud-users for running a specific streaming application. A paper has been submitted to an international conference in November 2017.

# 7.1.5. Stream Processing for Maritime Surveillance

Participants: Pascal Morillon, Christine Morin, Matthieu Simonin, Cédric Tedeschi.

In the context of maritime surveillance, and of the Sesame Project, we started the design and implementation of a platform dedicated to the batch and real-time processing of AIS messages sent by ships to inform about their identity, position and destination among other pieces of information.

Having use cases in mind such as detecting ships entering a protected areas, or ships having suspect behaviors, we designed a software architecture able to process AIS messages and produce synthetic data so as to answer these questions.

First experiments using a preliminary version of this platform have been conducted over the Grid'5000 platform using an archive of one-month of the AIS messages collected globally during March 2017. In particular, we've been able to index these messages using ElasticSearch<sup>0</sup> and visualize them using Kibana <sup>0</sup>.

The architecture has been described in a poster presented at BiDS'17 [56].

#### 7.1.6. Adaptive deployment for multi-cloud applications

Participants: Nikos Parlavantzas, Manh Linh Pham.

This work builds on the Adapter system, developed in the context of the PaaSage European project (2012-2016). The Adapter is part of the PaaSage open-source platform, a holistic solution for supporting the automatic deployment and execution of multi-cloud applications. Specifically, the Adapter is responsible for dynamic, cross-cloud application adaptation, taking into account adaptation costs and benefits in making deployment decisions. In 2017, we improved the Adapter and performed a comprehensive evaluation using experiments in a multi-cloud environment. The results demonstrate that Adapter supports automated multi-cloud adaptation while optimizing the performance and cost of the application. The results are described in an article currently under submission.

# 7.1.7. Application configuration and reconfiguration in multi-cloud environments

Participant: Nikos Parlavantzas.

Current approaches to cloud application configuration and reconfiguration are typically platform dependent, error prone and provide little support for optimizing application performance and resource utilisation. To address these limitations, we are combining the use of software product lines (SPLs) with performance prediction and automatic adaptation techniques. This work is performed in the context of the thesis of Carlos Ruiz Diaz, a PhD student at the University of Guadalajara, co-advised by Nikos Parlavantzas. The work has produced an SPL-based framework supporting initial configuration and dynamic adaptation in a systematic, platform-independent way.

In 2017, we extended the framework with a proactive adaptation solution that performs vertical VM scaling based on predictions of resource utilisation and performance. The solution targets multi-tier applications deployed on IaaS clouds. Experimental results demonstrate that the solution maintains expected application performance while reducing resource waste [46].

## 7.1.8. Adaptive resource management for high-performance, multi-sensor systems

Participants: Christine Morin, Nikos Parlavantzas, Baptiste Goupille-Lescar.

In the context of our collaboration with Thales Research and Technology and Baptiste Goupille-Lescar's PhD work, we are applying cloud resource management techniques to high-performance, multi-sensor, embedded systems with real-time constraints. The objective is to increase the flexibility and efficiency of resource allocation in such systems, enabling the execution of dynamic sets of applications with strict QoS requirements.

In 2017, we focused on an industrial use case concerning the operation of a multi-function surface active electronically scanned array (AESA) radar. We developed a simulation environment using an industrial high-precision AESA simulator and the Ptolemy II simulation framework, and we are using this environment to explore and evaluate different dynamic application placement solutions [57].

<sup>&</sup>lt;sup>0</sup>https://www.elastic.co/fr/

<sup>&</sup>lt;sup>0</sup>https://www.elastic.co/products/kibana

# 7.2. Greening Clouds

ICT (Information and Communications Technologies) ecosystem now approaches 6% of world electricity consumption and this ICT energy use will continue grow fast because of the information appetite of Big Data, big networks and big infrastructures as Clouds that unavoidably leads to big power.

# 7.2.1. Energy Models

Participants: Ehsan Ahvar, Loic Guegan, Anne-Cécile Orgerie, Martin Quinson.

Cloud computing allows users to outsource the computer resources required for their applications instead of using a local installation. It offers on-demand access to the resources through the Internet with a pay-as-you-go pricing model. However, this model hides the electricity cost of running these infrastructures.

The costs of current data centers are mostly driven by their energy consumption (specifically by the air conditioning, computing and networking infrastructure). Yet, current pricing models are usually static and rarely consider the facilities' energy consumption per user. The challenge is to provide a fair and predictable model to attribute the overall energy costs per virtual machine and to increase energy-awareness of users. We aim at proposing such energy cost models without heavily relying on physical wattmeters that may be costly to install and operate.

Another goal consists in better understanding the energy consumption of computing and networking resources of Clouds in order to provide energy cost models for the entire infrastructure including incentivizing cost models for both Cloud providers and energy suppliers. These models will be based on experimental measurement campaigns on heterogeneous devices. Inferring a cost model from energy measurements is an arduous task since simple models are not convincing, as shown in our previous work. We aim at proposing and validating energy cost models for the heterogeneous Cloud infrastructures in one hand, and the energy distribution grid on the other hand. These models will be integrated into simulation frameworks in order to validate our energy-efficient algorithms at larger scale.

# 7.2.2. Exploiting Renewable Energy in Clouds

Participants: Benjamin Camus, Yunbo Li, Anne-Cécile Orgerie.

The development of IoT (Internet of Things) equipment, the popularization of mobile devices, and emerging wearable devices bring new opportunities for context-aware applications in cloud computing environments. The disruptive potential impact of IoT relies on its pervasiveness: it should constitute an integrated heterogeneous system connecting an unprecedented number of physical objects to the Internet. Among the many challenges raised by IoT, one is currently getting particular attention: making computing resources easily accessible from the connected objects to process the huge amount of data streaming out of them.

While computation offloading to edge cloud infrastructures can be beneficial from a Quality of Service (QoS) point of view, from an energy perspective, it is relying on less energy-efficient resources than centralized Cloud data centers. On the other hand, with the increasing number of applications moving on to the cloud, it may become untenable to meet the increasing energy demand which is already reaching worrying levels. Edge nodes could help to alleviate slightly this energy consumption as they could offload data centers from their overwhelming power load and reduce data movement and network traffic. In particular, as edge cloud infrastructures are smaller in size than centralized data center, they can make a better use of renewable energy.

We propose to investigate the end-to-end energy consumption of IoT platforms. Our aim is to evaluate, on concrete use-cases, the benefits of edge computing platforms for IoT regarding energy consumption. We aim at proposing end-to-end energy models for estimating the consumption when offloading computation from the objects to the edge or to the core Cloud, depending on the number of devices and the desired application QoS, in particular trading-off between performance (response time) and reliability (service accuracy).

# 7.2.3. Smart Grids

Participants: Benjamin Camus, Anne-Cécile Orgerie, Martin Quinson.

We propose exploiting Smart Grid technologies to come to the rescue of energy-hungry Clouds. Unlike in traditional electrical distribution networks, where power can only be moved and scheduled in very limited ways, Smart Grids dynamically and effectively adapt supply to demand and limit electricity losses (currently 10% of produced energy is lost during transmission and distribution).

For instance, when a user submits a Cloud request (such as a Google search for instance), it is routed to a data center that processes it, computes the answer and sends it back to the user. Google owns several data centers spread across the world and for performance reasons, the center answering the user's request is more likely to be the one closest to the user. However, this data center may be less energy efficient. This request may have consumed less energy, or a different kind of energy (renewable or not), if it had been sent to this further data center. In this case, the response time would have been increased but maybe not noticeably: a different trade-off between quality of service (QoS) and energy-efficiency could have been adopted.

While Clouds come naturally to the rescue of Smart Grids for dealing with this big data issue, little attention has been paid to the benefits that Smart Grids could bring to distributed Clouds. To our knowledge, no previous work has exploited the Smart Grids potential to obtain and control the energy consumption of entire Cloud infrastructures from underlying facilities such as air conditioning equipment (which accounts for 30% to 50% of a data center's electricity bill) to network resources (which are often operated by several actors) and to computing resources (with their heterogeneity and distribution across multiple data centers). We aim at taking advantage of the opportunity brought by the Smart Grids to exploit renewable energy availability and to optimize energy management in distributed Clouds.

## 7.2.4. Involving Users in Energy Saving

Participants: David Guyon, Christine Morin, Anne-Cécile Orgerie.

In a Cloud moderately loaded, some servers may be turned off when not used for energy saving purpose. Cloud providers can apply resource management strategies to favor idle servers. Some of the existing solutions propose mechanisms to optimize VM scheduling in the Cloud. A common solution is to consolidate the mapping of the VMs in the Cloud by grouping them in a fewer number of servers. The unused servers can then be turned off in order to lower the global electricity consumption.

Indeed, current work focuses on possible levers at the virtual machine suppliers and/or services. However, users are not involved in the choice of using these levers while significant energy savings could be achieved with their help. For example, they might agree to delay slightly the calculation of the response to their applications on the Cloud or accept that it is supported by a remote data center, to save energy or wait for the availability of renewable energy. The VMs are black boxes from the Cloud provider point of view. So, the user is the only one to know the applications running on her VMs.

We plan to explore possible collaborations between virtual machine suppliers, service providers and users of Clouds in order to provide users with ways of participating in the reduction of the Clouds energy consumption. This work will follow two directions: 1) to investigate compromises between power and performance/service quality that cloud providers can offer to their users and to propose them a variety of options adapted to their workload; and 2) to develop mechanisms for each layer of the Cloud software stack to provide users with a quantification of the energy consumed by each of their options as an incentive to become greener.

Our results were published in [40], [32], [31].

# 7.3. Securing Clouds

#### 7.3.1. Security Monitoring in Clouds

**Participants:** Christine Morin, Jean-Louis Pazat, Louis Rilling, Anna Giannakou, Amir Teshome Wonjiga, Clément El Baz.

In the INDIC project we aim at making security monitoring a dependable service for IaaS cloud customers. To this end, we study three topics:

- defining relevant SLA terms for security monitoring,
- enforcing and verifying SLA terms,
- making the SLA terms enforcement mechanisms self-adaptable to cope with the dynamic nature of clouds.

The considered enforcement and verification mechanisms should have a minimal impact on performance.

In 2017, we did a thorough performance evaluation and security correctness analysis of the SAIDS approach, that we proposed in 2015, and that makes a network intrusion detection system (NIDS) deployed in a cloud operator infrastructure self-adaptable. In the performance evaluation we studied the performance impact of SAIDS on the cloud infrastructure operations related to the management of virtual machines (typically creation, migration, and deletion) as well as the scalability of SAIDS with respect to the number of NIDS devices managed. This performance evaluation was done on the Grid'5000 platform. The results showed that SAIDS adds very low overhead and is scalable. The security analysis was done both experimentally and based on a risk analysis. This analysis validated the security correctness of SAIDS. A full paper presenting SAIDS and its evaluation is submitted for publication in 2018. A demo of SAIDS was presented at FIC 2017, Lille, France in January 2017 and at the Inria Industry Days, Paris, France on October 17th, 2017.

Regarding SLA definition and enforcement, in 2017 we evaluated the verification method that we defined in 2016 and that enables a Cloud customer to verify that an NIDS located in the operator infrastructure is configured correctly according to the Service-Level Objectives (SLO) figuring in the SLA. The performance evaluation was done on the Grid'5000 platform and showed that the proposed verification method requires making a trade-off between verification speed and impact on the performance of the production applications deployed in the tenant's virtual machines. The security correctness analysis was based on a risk analysis and showed the constraints on the types of attacks that can be used for verification as well as the limitations due to the tools used in the prototype [55]. A full paper presenting the verification method and its evaluation is submitted for publication in 2018.

After the acquired experience on verifying security monitoring metrics, we started studying how to define relevant SLOs that are verifiable. We plan to get results in 2018 and submit a paper for publication in 2018 or 2019.

Finally, in October 2017 we started studying how security monitoring SLAs could take into account context changes like the evolution of threats and updates to the tenants' software.

Our work done as part of the INDIC project were presented in [59].

#### 7.3.2. Risk assessment in clouds

#### Participant: Christine Morin.

Cloud providers have an incomplete view of their hosted virtual infrastructures managed by a Cloud Management System (CMS) and a Software Defined Network (SDN) controller. For various security reasons (e.g. isolation verification, modeling attack paths in the network), it is necessary to know which virtual machines can interact via network protocols. This requires building a connectivity graph between the virtual machines, that we can extract with the knowledge of the overall topology and the deployed network security policy. Existing methodologies for building such models for physical networks produce incomplete results. Moreover, they are not suitable for cloud infrastructures due to either their intrusiveness or lack of connectivity discovery. We propose a method to compute the connectivity graph, relying on information provided by both the CMS and the SDN controller. Connectivity can first be extracted from knowledge databases, then dynamically updated on the occurrence of cloud-related events. We realized an experimental evaluation of the proposed method to determine its correctness and performance in a realistic context, considering CPU and RAM consumption, the volume of data generated, and execution time for the different portions of the algorithm involved. Experiments were run on the Grid'5000 platform with OpenStack CMS and ONOS SDN controller. Our approach proves on a representative infrastructure to compute exact, complete and up-to- date connectivity graphs in reasonable time [42], [41].

#### 7.3.3. Personal Data Management in Cloud-based IoT Systems

Participants: Christine Morin, Jean-Pierre Banâtre, Deborah Agarwal, Subhadeep Sarkar, Louis Rilling.

The Internet of Things (IoT), in today's digital world, encompasses billions of smart connected devices. These devices generate an unprecedented amount of data, which often bears sensitive personal information of individuals. In present service models, the data are processed and managed by service providers, beyond the visibility of the owner of the data. Although the EU General Data Protection Regulation (GDPR) strives to protect citizens and their data by regulation, citizens and service providers need technological advances to gain effective control over their data or to prove compliance with the new regulation. Our primary objective is to enforce, by design, the GDPR at the system level so as to preserve the privacy concerning personal data. We started off with enforcement of the data erasure facility as expressed in the GDPR. Data erasure corresponds to both automatic erasure of data after expiration of their retention period and ad-hoc on request of the data owner. Our first contribution, towards this, is design of a customizable privacy policy, which would allow the end users to express their preferences regarding the purpose of use, location of processing, retention period, sharing and storage policies concerning their personal data. We developed a XML-based policy expression language by defining the required data structures and vocabulary, which will facilitate the end-users to easily express their preferences. Next, we have investigated into the possible way of the implementation of the proposed solution and identified the exploitation of the operation system capabilities as an appropriate means to the cause. For this, we have potentially chosen the Sel4 (or may be some other capability-based microkernel) as our platform of operation. Finally, we have identified the different challenges towards implementation of our solution and did some groundwork towards proposing the solutions to the same. These challenges include efficient identification of replication of data, locating all replicas of a given data segment, and implementing erasure of data in a cross-domain service model.

# **7.4. Experimenting with Clouds**

# 7.4.1. Simulation

Participants: Martin Quinson, Loic Guegan, Toufik Boubehziz, The Anh Pham.

We propose to combine two complementary experimental approaches: direct execution on testbeds such as Grid'5000, that are eminently believable but rather labor intensive, and simulations (using *e.g.* SimGrid) that are much more light-weighted, but requires are careful assessment. One specificity of the Myriads team is that we are working on these experimental methodologies *per se*, raising the standards of *good experiments* in our community. The Grid'5000 operational team is embedded in our research team, ensuring that our work remains aligned with the ground reality.

In 2017, our work was mostly centered on letting SimGrid become a *de facto* standard for the simulation of distributed platforms. We introduced a new programming interface, particularly adapted to the study of abstract algorithms. Beyond the engineering task, this requires to carefully capture the concepts that are important to the practitioners on distributed systems.

SimGrid is not limited to abstract algorithms, and can also be used to simulate real applications. This year, we published a journal article on the many challenges to overcome when designing a simulator of high performance systems. This work was published in the TPDS journal [20].

On the modeling side, our team worked this year toward the improvement of energy models, both for computational facilities and for the network. Despite the scarce availability of real testbeds that allow fine-grained energy measurements, we managed to provide a generic energy consumption model, published in [35], [43].

Finally, we restarted our efforts toward the formal verification of distributed systems. The model-checker that is integrated within SimGrid is already functional ([44]), but more work is necessary to make it efficient. We even found cases for which our reduction algorithm may miss defects in the verified system. This work will certainly motivate much more work in the future years.

## 7.4.2. Use cases

Participants: Christine Morin, Nikos Parlavantzas, Deborah Agarwal, Manh Linh Pham.

## 7.4.2.1. Simulation framework for studying between-herd pathogen spread in a region

In the context of the MIHMES project (2012-2017) and in collaboration with INRA researchers, we transformed a legacy application for simulating the spread of bovine viral diarrhea virus (BVDV) to a cloudenabled application based on the DiFFuSE framework (Distributed framework for cloud-based epidemic simulations). Specifically, the original sequential code was first modified to add single-computer parallelism using OpenMP. We then decomposed the code into separate services that were deployed across multiple clouds and independently scaled. Using this service-based cloud-enabled simulation, we performed a set of experiments that demonstrated that applying DiFFuSE increases performance, allows exploring different cost-performance trade-offs, automatically handles failures, and supports elastic allocation of resources from multiple clouds [45].

#### 7.4.2.2. FluxNet and AmeriFlux Data Analysis

The carbon flux datasets from AmeriFlux (Americas) and FLUXNET (global) are comprised of long-term time series data and other measurements at each tower site. There are over 800 flux towers around the world collecting this data. The non-time series measurements include information critical to performing analysis on the site's data. Examples include: canopy height, species distribution, soil properties, leaf area, instrument heights, etc. These measurements are reported as a variable group where the value plus information such as method of measurement and other information are reported together. Each variable group has a different number and type of parameters that are reported. The current output format is a normalized file. Users have found this file difficult to use.

Our earlier work in the DALHIS Inria associate team focused on building user interfaces to specify the data. This year we jointly worked on developing a Jupyter Notebook that would serve as a tool for users to read in and explore the data in a personalized tutorial type environment. We developed two notebooks and the next step is to start user testing on the notebooks.

# 8. Bilateral Contracts and Grants with Industry

# 8.1. Bilateral Contracts with Industry

# 8.1.1. Technicolor (2016-2017)

## Participant: Guillaume Pierre.

Our collaboration with Technicolor has focused on the design of a scalable and elastic virtual customer premises equipment based on Network Function Virtualization, Software-Defined Networking and Cloud technologies. In 2017 we completed the system design and started implementing the system. The collaboration completed successfully in June 2017. However, the vCPE project within which this collaboration took place was unfortunately interrupted by Technicolor before we could write a paper about this work.

# **8.2.** Bilateral Grants with Industry

## 8.2.1. Thales Research and Technology (2016-2018)

Participants: Baptiste Goupille-Lescar, Christine Morin, Nikos Parlavantzas.

Our collaboration with Thales Research and Technology focuses on the development of distributed Cyber-Physical Systems, such as those developed by Thales to monitor and react to changing physical environments. These systems need to be highly adaptable in order to cope with the dynamism and diversity of their operating environments. Notably, they require distributed, parallel architectures that support dynamic sets of applications, not known in advance, while providing strong QoS guarantees. The objective of this collaboration is to explore adaptive resource management mechanisms for such systems that can adapt to changes in the requirements and in the availability of resources. This contract funds Baptiste Goupille-Lescar's PhD grant.

## 8.2.2. Nokia (2015-2018)

Participant: Christine Morin.

Together with CIDRE Inria project-team we are involved in a collaboration with Nokia on security policy adaptation driven by risk evaluation in modern communication infrastructures. To address the need for efficient security supervision mechanisms, approaches such as attack graphs generation, coupled to a risk-based assessment have been used to provide an insight into a system's threat exposure. In comparison to static infrastructures, clouds exhibit a dynamic nature and are exposed to new attack scenarios due to virtualization. The goal of this collaboration is thus to revisit existing methods in the context of clouds. This contract funds Pernelle Mensah's PhD grant. Pernelle is a member of CIDRE project-team.

# 9. Partnerships and Cooperations

# 9.1. Regional Initiatives

#### 9.1.1. CominLabs EPOC project (2013-2017)

Participants: Sabbir Hasan Rochi, Yunbo Li, Anne-Cécile Orgerie, Jean-Louis Pazat.

In this project, partners aim at focusing on energy-aware task execution from the hardware to application's components in the context of a mono-site data center (all resources are in the same physical location) which is connected to the regular electric Grid and to renewable energy sources (such as windmills or solar cells). In this context, we tackle three major challenges:

- Optimizing the energy consumption of distributed infrastructures and service compositions in the presence of ever more dynamic service applications and ever more stringent availability requirements for services.
- Designing a clever cloud's resource management which takes advantage of renewable energy availability to perform opportunistic tasks, then exploring the trade-off between energy saving and performance aspects in large-scale distributed systems.
- Investigating energy-aware optical ultra high-speed interconnection networks to exchange large volumes of data (VM memory and storage) over very short periods of time.

Sabbir Hasan Rochi has defended his PhD on SLA driven Cloud autoscaling for optimizing energy footprint on May 3rd, 2017. Yunbo Li has defended his PhD on Resource allocation in a Cloud partially powered by renewable sources on June 12th, 2017.

# 9.1.2. INDIC - Cybersecurity Pole of Excellence (2014-2020)

**Participants:** Anna Giannakou, Christine Morin, Jean-Louis Pazat, Louis Rilling, Amir Teshome Wonjiga, Clément El Baz.

Our study carried out in the framework of a collaboration with DGA-MI aims at defining and enforcing SLA for security monitoring of virtualized information systems. To this aim we study three topics:

- defining relevant SLA terms for security monitoring,
- enforcing and evaluating SLA terms,
- making the SLA terms enforcement mechanisms self-adaptable to cope with the dynamic nature of clouds.

The considered enforcement and evaluation mechanisms should have a minimal impact on performance. The funding from DGA funds two PhD students: Anna Giannakou and Amir Teshome Wonjiga. Clément El Baz is partially funded by the Brittany Regional Council in the PEC framework.

# 9.2. National Initiatives

## 9.2.1. ADEME RennesGrid

Participants: Benjamin Camus, Anne-Cécile Orgerie, Martin Quinson.

The aim of the RennesGrid project is to design and implement a large-scale preindustrial microgrid demonstrator in the territory of Rennes Metropole to organize the shared self-consumption of a group of photovoltaic panels coupled to stationary storage devices. Traditional approaches to power grid management tend to overlook the costs, both energy and economic, of using computers to ensure optimal electricity network management. However, these costs can be significant. It is therefore necessary to take them into account along with the design of IT tools during studies of optimal energy management of smart grids. In addition, telecommunication networks are generally considered to have an ideal functioning, that is to say they can not negatively affect the performance of the electricity network. However, this is not realistic and it is necessary to analyze the impact of phenomena such as congestion, latency, failures related to computer equipment or impact on the batteries of sensors, etc. on strategies for optimal management of the electricity network. In this project, we will closely collaborate with Anne Blavette (CR CNRS in electrical engineering, SATIE, Rennes) and co-supervise a post-doc on evaluating the impact of the IT infrastructure in the management of smart grids.

# 9.2.2. Inria ADT SaaP (2016-2018)

Participants: Toufik Boubehziz, Martin Quinson.

The SaaP technological development action (SimGrid As A Platform) funded by INRIA targets the refactoring of SimGrid to make it ready to use in production and teaching contexts. Our ultimate goal is to sustain the development of the framework by involving 5 to 10 companies that are using it internally. Our target of the teaching context is thus an intermediate goal, as we think that the best solution to ensure the adoption of our tool by the industrial engineers is that they discover the tool during their studies.

The technical actions envisioned for this ADT are the complete re-factoring of the software (to make it easier to script a new model within the tool kernel) and a reorganization of the interfaces (for a better integration in the Java and python language). This work is lead by Toufik Boubehziz in collaboration with the whole SimGrid community, which provide valuable feedback.

## 9.2.3. Inria ADT DiFFuSE (2017-2018)

Participants: Nikos Parlavantzas, Christine Morin, Manh Linh Pham.

The DiFFuSE technological development action (Distributed framework for cloud-based epidemic simulations) funded by INRIA focuses on the DiFFuSE framework developed by Myriads in the context of MIHMES (2012-2017). MIHMES was a 5-year collaborative multidisciplinary project funded by ANR under the Investments for the Future Program, and led by BIOEPAR, INRA, ONIRIS. DiFFuSE is a framework that provides design support, reusable code, and tools for building and executing epidemic simulations in the cloud. The main objectives of this ADT are to improve the usability and robustness of DiFFuSE, to provide support to scientists for applying the framework to a new epidemic simulations as well as to provide a thorough evaluation of the framework using multiple case studies.

# 9.2.4. Inria IPL Discovery (2015-2019)

Participants: Ehsan Ahvar, Anne-Cécile Orgerie, Matthieu Simonin, Genc Tato, Cédric Tedeschi.

The Inria IPL Discovery officially started in September 2015. It targets the design, development and deployment of a distributed Cloud infrastructure within the network's backbone. It will be based upon a set of building blocks whose design will take locality as a primary constraint, so as to minimize distant communications and consequently achieve better network traffic, partition management and improved availability.

Its developments are planned to get integrated within the OpenStack framework. Myriads is involved in the design of new overlay networks for such environments so as to support efficient messaging and routing. Myriads is also involved in the energy/cost benefit analysis of distributed edge-cloud architectures.

# 9.2.5. Inria IPL CityLab (2015-2018)

Participants: Subarna Chatterjee, Christine Morin.

The Inria Project Lab (IPL) CityLab@Inria (http://citylab.inria.fr) studies ICT solutions toward smart cities that promote both social and environmental sustainability. A strong emphasis of the Lab is on the undertaking of a multi-disciplinary research program through the integration of relevant scientific and technology studies, from sensing up to analytics and advanced applications, so as to actually enact the foreseen smart city Systems of Systems. City-scale experiments of the proposed platforms and services are planned in cities in California and France, thereby learning lessons from diverse setups.

Myriads investigates advanced cloud solutions for the Future Internet, which are critical for the processing of urban data. It leverages its experience in cloud computing and Internet of services while expanding its research activities to the design and implementation of cloud services to support crowd-Xing applications and mobile social applications.

In 2017, Christine Morin was involved in the preparation of a SPOC entitled "Technological challenges of participatory smart cities", which is proposed in the framework of the EIT Digital professional school. She prepared seven sequences on cloud-based urban data management. This SPOC is the English version of the MOOC entitled "*Défis technologiques des villes intelligentes participatives*" run on the FUN platform in Spring and Fall 2017.

In 2017, we also conducted a comparative experimental evaluation of data stream processing environments executed on clusters and clouds. We compared the performance and energy consumption of Heron, Storm and Flink frameworks with three data streaming representative applications.

#### 9.2.6. Inria IPL Hac Specis (2016-2020)

Participants: Anne-Cécile Orgerie, Martin Quinson, The Anh Pham.

The goal of the HAC SPECIS (High-performance Application and Computers: Studying PErformance and Correctness In Simulation) project (http://hacspecis.gforge.inria.fr/) is to answer methodological needs of HPC application and runtime developers and to allow to study real HPC systems both from the correctness and performance point of view. To this end, we gather experts from the HPC, formal verification and performance evaluation community.

During his first year of PhD thesis, The Anh Pham conducted an analysis of the formal methods and algorithms used in SimGrid. This work, co-advised by Martin Quinson with Thierry Jéron (team SUMO, formal methods), was important to bridge the gap between the involved communities. The resulting work has been published in a workshop gathering the intersection between the communities of formal methods and HPC [44].

Another PhD thesis will stat in December 2017, co-advised by Laurent Lefèvre (Avalon team, Lyon), Martin Quinson and Anne-Cécile Orgerie. This thesis will focus on simulating the energy consumption of continuum computing between heterogeneous numerical infrastructures for HPC.

# 9.2.7. COSMIC PRE (2016 - 2018)

Participants: Benjamin Camus, Anne-Cécile Orgerie, Martin Quinson.

The distributed nature of Cloud infrastructures involves that their components are spread across wide areas, interconnected through different networks, and powered by diverse energy sources and providers, making overall energy monitoring and optimization challenging. The COSMIC project aims at taking advantage of the opportunity brought by the Smart Grids to exploit renewable energy availability and to optimize energy management in distributed Clouds. This PRE, led by Anne-Cécile Orgerie also involves Fanny Dufossé from Dolphin team (Inria Lille), Anne Blavette from SATIE laboratory (electrical engineering, Rennes), and Benjamin Camus, who has started a 18 months post-doc in October 2016 in the context of this project. A paper on this project has been presented at SMARTGREENS 2017 and two others are currently under submission.

# 9.2.8. MIHMES ANR Investissements d'Avenir (2012 - 2017)

Participants: Christine Morin, Manh Linh Pham, Nikos Parlavantzas.

The MIMHES project (http://www.inra.fr/mihmes) led by INRA/BioEpAR aimed at producing scientific knowledge and methods for the management of endemic infectious animal diseases and veterinary public health risks. The role of Myriads was to help MIHMES researchers improve the performance of their simulation applications and take advantage of computing resources provided by clouds. To that end, Myriads developed a framework, named DiFFuSE, that provides design support, reusable code, and tools for building and executing epidemic simulations in the cloud.

In 2017, we further developed DiFFuSE and extended the framework to make use of the PaaSage open-source platform, the main outcome of a European FP7 IP project in which Myriads participated (2012-2016). Thanks to PaaSage, DiFFuSE allows deploying and managing services in multi-cloud environments. We applied DiFFuSE to restructure an application that simulates the spread of the bovine viral diarrhea virus (BVDV) and conducted experiments to evaluate DiFFuSE [45].

# 9.2.9. SESAME ASTRID project (2016-2019)

Participants: Pascal Morillon, Christine Morin, Matthieu Simonin, Cédric Tedeschi, Mehdi Belkhiria.

The Sesame project (http://www.agence-nationale-recherche.fr/Project-ANR-16-ASTR-0026) led by IMT Atlantique aims at develop efficient infrastructures and tools for the maritime traffic surveillance. The role of Myriads is to define a robust and scalable infrastructure for the real-time and batch processing of vessel tracking information.

## 9.2.10. PIA ELCI (2015-2018)

Participant: Anne-Cécile Orgerie.

The PIA ELCI project deals with software environment for computation-intensive applications. It is leaded by BULL. In the context of this project, we collaborate with ROMA and Avalon teams from Lyon: we cosupervise a PhD student (Issam Rais) funded by this project with these teams on multi-criteria scheduling for large-scale HPC environments. This collaboration has led to two publications in 2017: a journal article published in IJHPCA and a conference paper presented at EuroPar.

## 9.2.11. CNRS GDS EcoInfo

Participant: Anne-Cécile Orgerie.

The EcoInfo group deals with reducing environmental and societal impacts of Information and Communications Technologies from hardware to software aspects. This group aims at providing critical studies, lifecycle analyses and best practices in order to improve the energy efficiency of printers, servers, data centers, and any ICT equipment in use in public research organizations. In particular, it has led in December 2016 to the publication of an ADEME report jointly with Deloitte Développement Durable, Futuribles, CREDOC and ADEME on the potential contribution of digital to the reduction of environmental impacts: state of play and challenges for the prospective.

# 9.3. European Initiatives

## 9.3.1. FP7 & H2020 Projects

#### 9.3.1.1. FogGuru

Participant: Guillaume Pierre.

Title: MSCA ITN EID Programm: H2020 Duration: September 2017 - August 2021 Coordinator: Guillaume Pierre Participants:

University of Rennes 1, France (coordinator) Technisch Universität Berliun, Germany Elastisys AB, Sweden U-Hopper srl, Italy EIT Digital Rennes, France Las Naves, Spain

FogGuru is a doctoral training project which aims to to train eight talented PhD students with an innovative and inter-sectoral research program to constitute the next generation of European Cloud and Fog computing experts. Besides their scientific and technical education, FogGuru's PhD students will receive extensive training in technological innovation and entrepreneurship as well as soft skills. These combined skills will enable them to fully master the innovative products and services, and to real-life deployment, experimentation and engagement with beta-testers.

#### 9.3.1.2. Fed4Fire+ (2017-2022)

Participants: David Margery, Yue Li.

Title: Federation for FIRE Plus

Programm: H2020

Duration: January 2017 - December 2021

Coordinator: Interuniversitair Micro-Electronicacentrum Imec VZW

Partners:

Universidad de Malaga

National Technical University of Athens - NTUA

The Provost, Fellows, Foundation Scholars & the other members of board of the College of the Holy & Undivided Trinity of Queen Elizabeth Near Dublin

Ethniko Kentro Erevnas Kai Technologikis Anaptyxis

GEANT LImited

Institut Jozef Stefan

Mandat International Alias Fondation Pour la Cooperation Internationale

Université Pierre et Marie Curie - Paris 6

Universidad De Cantabria

Fundacio Privada I2CAT, Internet I Innovacio Digital A Catalunya

EURESCOM-European Institute For Research And Strategic Studies in Telecommunications GMBH

Nordunet A/S

Technische Universitaet Berlin

Instytut Chemii Bioorganicznej Polskiej Akademii Nauk

Fraunhofer Gesellschaft zur Foerderung Der Angewandten Forschung E.V.

Universiteit Van Amsterdam

University of Southampton

Martel GMBH

Atos Spain SA

Institut National de Recherche en Informatique et automatique

#### Inria contact: David Margery

Fed4FIRE+ is a successor project to Fed4FIRE. In Fed4FIRE+, we more directly integrate Grid'5000 into the wider eco-system of experimental platforms in Europe and beyond using results we developed in Fed4FIRE. We have developed a generalized proxy mechanisms to allow users with Fed4FIRE identities to interact with services giving access to different testbeds but not designed to support Fed4FIRE identities. Fed4FIRE+ has started January 1st, 2017.

#### 9.3.2. Collaborations in European Programs, Except FP7 & H2020

#### 9.3.2.1. NESUS

Participant: Anne-Cécile Orgerie.

Program: ICT COST

Project acronym: NESUS

Project title: Network for Sustainable Ultrascale Computing (ICT COST Action IC1305)

Duration: 2014 - 2018

Coordinator: Prof. Jesus Carretero, University Carlos III of Madrid, Spain, http://www.nesus.eu

Other partners: 33 COST countries and 11 non-COST countries

Abstract: Ultrascale systems are envisioned as large-scale complex systems joining parallel and distributed computing systems that will be two to three orders of magnitude larger that today's systems. The EU is already funding large scale computing systems research, but it is not coordinated across researchers, leading to duplications and inefficiencies. The goal of the NESUS Action is to establish an open European research network targeting sustainable solutions for ultrascale computing aiming at cross fertilization among HPC, large scale distributed systems, and big data management. The network will contribute to gluing disparate researchers working across different areas and provide a meeting ground for researchers in these separate areas to exchange ideas, to identify synergies, and to pursue common activities in research topics such as sustainable software solutions (applications and system software stack), data management, energy efficiency, and resilience. Some of the most active research groups of the world in this area are members of this proposal. This Action will increase the value of these groups at the European-level by reducing duplication of efforts and providing a more holistic view to all researchers, it will promote the leadership of Europe, and it will increase their impact on science, economy, and society. Anne-Cécile Orgerie is co-responsible of the focus group on metrics, monitoring, instrumentation and profiling in the Working Group 5 on Energy Efficiency. A joint paper has been accepted in 2017 on this topic at the Elsevier journal on Sustainable Computing.

#### 9.3.3. Collaborations with Major European Organizations

Partner 1: EPFL, Network architecture lab (Switzerland)

We collaborate with Katerina Argyraki's research group on the integration of networking and cloud computing technologies in order to support placement constraints between cloud resources.

Partner 2: University of Neuchâtel, dept. of Computer Science (Switzerland)

We collaborate with Pascal Felber's research group on energy efficiency in Clouds and in particular on the design of energy cost models for virtual machines. A joint journal paper has been accepted in 2017 for publication in Sustainable Computing: Informatics and Systems, Elsevier.

Partner 3: Catholic University of Louvain (Belgium)

We collaborate with Etienne Riviere's research group on the efficient service placement and discovery in a SaaS context.

# 9.4. International Initiatives

# 9.4.1. Inria International Labs

#### 9.4.1.1. DALHIS

**Participants:** Christine Morin, Deborah Agarwal, Anna Giannakou, Amir Teshome Wonjiga, Subarna Chatterjee.

Title: Data Analysis on Large Heterogeneous Infrastructures for Science

International Partner (Institution - Laboratory - Researcher):

Lawrence Berkeley National Laboratory (United States) - Data Science and Technology department - Deb Agarwal

Start year: 2016

See also: https://project.inria.fr/dalhis/

Data produced by scientific instruments (large facilities like telescopes or field data), large-scale experiments, and high-fidelity simulations are increasing in magnitude and complexity. Existing data analysis methods, tools and infrastructure are often difficult to use and unable to provide the complete data management, collaboration, and curation environment needed to manage these complex, dynamic, and large-scale data analysis environments. The goal of the Inria-LBL DALHIS associate team involving the Myriads (PI) and Avalon Inria project-teams and the Data Science and Technology (DST) department at Lawrence Berkeley National Laboratory (LBL) is to create a collaborative distributed software ecosystem to manage data lifecycle and enable data analytics on distributed data sets and resources. Specifically, our goal is to build a dynamic software stack that is user-friendly, scalable, energy-efficient and fault tolerant. Our research will determine appropriate execution environments that allow users to seamlessly execute their end-to-end dynamic data analysis workflows in various resource environments and scales while meeting energy-efficiency, performance and fault tolerance goals. We will engage in deep partnerships with scientific teams (Fluxnet in environmental science and SNFactory and LSST experiences in cosmology) and use a mix of user research with system software R&D to address specific challenges that these communities face. In 2017, we worked on evaluating data streaming environments (see Section 7.1.4) and on producing tools to help users (scientist in the climate and environment community) to explore the carbon flux datasets from AmeriFlux (Americas) and FLUXNET (global) (see Section 7.4.2.2). We also worked on two facets of security in the context of HPC distributed computing infrastructures: (i) building a workflow for data analysis for anomaly detection and (ii) using the block-chain technology to leverage data integrity at the network and some portion of computation levels.

#### 9.4.2. Inria Associate Teams Not Involved in an Inria International Labs

#### 9.4.2.1. SUSTAM associated team

Participants: Anne-Cécile Orgerie, Yunbo Li.

Anne-Cécile Orgerie participates in the associated team named SUSTAM (Sustainable Ultra Scale compuTing, dAta and energy Management) leaded by Laurent Lefévre (Avalon team, Lyon) with Prof. Manish Parashar (RDI2, Rutgers University, NJ, USA). The SUSTAM associate team will focus on the joint design of a multicriteria orchestration framework dealing with resources, data and energy management in an sustainable way.

### 9.4.3. Inria International Partners

#### 9.4.3.1. Informal International Partners

Partner: Rutgers University, dept. of Computer Science (New Jersey, United States)

We collaborate with Manish Parashar's research group on energy efficiency in edge Clouds and in particular on the design of energy cost models for such environments involving renewable energy. A joint paper has been presented at IEEE/ACM CCGrid 2017.

Partner: Northeastern University, dept. of Computer Science (Massachusetts, United States)

We collaborate with Gene Cooperman 's research group on the study of large-scale distributed systems. More specifically, we actively collaborate on virtualization technologies and system snapshoting (we obtained a postdoc funding on that topic from the Brittany Regional Council, but all applicants declined in the last minute). We plan to reinforce and extend our collaboration to formal methods for distributed systems in the next year.

Partner: University of Guadalajara (Mexico)

We collaborate with the team of Prof. Hector Duran-Limon on application and resource management in the cloud. In 2017, we produced a joint publication [46]. Nikos Parlavantzas is co-advising a PhD student enrolled in the University of Guadalajara (Carlos Ruiz Diaz).

Partner: Tlemcen University (Algeria)

We collaborate with Djawida Dib on energy-efficient fault-tolerant resource and application management in containerized clouds. Christine Morin and Nikos Parlavantzas have been co-advising Yasmina Bouizem, a PhD student enrolled in the University of Tlemcen from December 2016.

# 9.5. International Research Visitors

## 9.5.1. Visits of International Scientists

Deb Agarwal, senior scientist at Lawrence Berkeley National Laboratory, who has been awarded an Inria International Chair for the 2015-2019 period, visited Myriads team during two months from August 1st to October 15th, 2017.

Professor Gene Cooperman, Northeastern University, Boston, USA, visited the Myriads team for one week in July to reinforce our collaboration on the virtualization of large-scale distributed systems.

#### 9.5.1.1. Internships

Betsegaw Lemma Amersho Date: Feb-June 2017 Institution: University of Rennes 1 & Aalto University (Finland) Supervisors: Anne-Cécile Orgerie and Martin Quinson Vinothkumar Nagasayanan Date: May-August 2017 Institution: University of Rennes 1 & TU Berlin (Germany) Supervisor: Guillaume Pierre Salsabil Amri Date: May-August 2017 Institution: University of Rennes 1 Supervisor: Guillaume Pierre Bérenger Nguyen Nhon Date: July-August 2017 Institution: University of Rennes 1 Supervisor: Guillaume Pierre

# 9.5.2. Visits to International Teams

#### 9.5.2.1. Research Stays Abroad

• Genc Tato started a 6-month research visit at the Catholic University of Louvain in November 2017 to work with Etienne Riviere on service placement and discovery in a SaaS context.

- Amir Teshome Wonjiga did a 3-month research internship in the Data Science and Technology department of the Lawrence Berkeley National Laboratory from April to June 2017. He worked with Sean Peisert, staff scientist, on ensuring data integrity in the workflow of high performance applications.
- Anne-Cécile Orgerie visited for 1 week the team of Prof. Manish Parashar in the RDI2 laboratory at Rutgers University in October 2017.

# **10.** Dissemination

# **10.1. Promoting Scientific Activities**

# 10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- Martin Quinson was the main organizer of the research school SUD gathering the SimGrid user community, held on 20-24 November, in Le Bono, Brittany.
- Guillaume Pierre was a co-organizer of the NetV IRISA/Technicolor Workshop on Network Virtualization (Rennes, February 1st 2017)
- 10.1.1.2. Member of the Organizing Committees

Guillaume Pierre is general chair for the ACM/IFIP/Usenix Middleware 2018 conference. Christine Morin and Benjamin Camus are respectively sponsor and publicity & web chairs for this conference.

# 10.1.2. Scientific Events Selection

- 10.1.2.1. Chair of Conference Program Committees
  - Anne-Cécile Orgerie was Program vice co-chair for GreenCom 2017: IEEE International Conference on Green Computing and Communications.
  - Anne-Cécile Orgerie was co-chair of the Track on Performance Modeling and Evaluation for ICA3PP 2017: International Conference on Algorithms and Architectures for Parallel Processing.
- 10.1.2.2. Member of the Conference Program Committees
  - Christine Morin served as program committee member for the following conferences: ACM International Conference on High Performance Distributed Systems (HPDC) 2017, IEEE International Conference on Distributed Computing Systems (ICDCS) 2017, ACM/IEEE SC 2017, IEEE Big Data 2017, IEEE CloudCom 2017, IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CC-GRID) 2017, IEEE Conference on Network unction Virtualization and Software Defined Networks (NFV-SDN) 2017, ComPas 2017 and workshop: Resilience 2017.
  - Anne-Cécile Orgerie served as a program committee member for the following conferences: PDP 2017, ICC 2017, I-SPAN 2017, Compas 2017, CBDCom 2017, SC'17, SD3C 2017, and ISPA 2017.
  - Guillaume Pierre served as a program committee member of ACM SAC 2017, ISFEC 2017, ISPDC 2017, CloudCom 2017, and HotWeb 2017.
  - Nikos Parlavantzas served as a program committee member of ISPDC 2017, VHPC 17, and CLOSER 2017.
  - Cédric Tedeschi served as a program committee member for ICCS 2017, Closer 2017 and Compas 2017.

# 10.1.3. Journal

#### 10.1.3.1. Member of the Editorial Boards

Christine Morin is associate editor in the IEEE's Transactions on Parallel and Distributed Systems' Editorial Board.

#### 10.1.3.2. Reviewer - Reviewing Activities

- Anne-Cécile Orgerie served as a reviewer for SIMPAT Simulation Modelling Practice and Theory in 2017.
- Cédric Tedeschi served as a reviewer for TPDS in 2017.

# 10.1.4. Invited Talks

- Anne-Cécile Orgerie gave a talk for the seminar organized at CEA *maison de la simulation* in December 2017 on modeling and simulating the energy consumption of HPC infrastructures.
- Anne-Cécile Orgerie gave a talk for the Inria Scientific Days in June 2017 on measuring the energy consumption and improving the energy efficiency of computing infrastructures.
- Anne-Cécile Orgerie gave a talk for a seminar organized by ENS Rennes for high school students in March 2017 on greening ICT.
- Anne-Cécile Orgerie gave a talk for a seminar organized by the Seine-et-Marne DSDEN during a training day for teachers of primary, secondary and high schools on ICT and sustainability in March 2017 on the energy consumption of ICT.
- Anne-Cécile Orgerie gave a talk for a public conference organized by ArmorScience, an association based in Lannion, in February 2017 on reducing the energy impacts of ICT.
- Anne-Cécile Orgerie gave a talk for a seminar at ENSSAT in Lannion in the Informatics department in February 2017 on saving energy in large-scale distributed systems.
- Anne-Cécile Orgerie gave a talk for the IRISA-Technicolor Workshop on Network Virtualization on February 2017 on greening the networks.
- Guillaume Pierre gave invited talks at the RESCOM summer school on virtualization and containerization (Le Croisic, France, June 2017), the 11th Cloud Control Workshop (Sweden, June 2017), in a seminar of the DIMA team at TU Berlin (November 2017), and a guest lecture at INSA Rennes (France, November 2017).

# 10.1.5. Leadership within the Scientific Community

- Anne-Cécile Orgerie is co-responsible for the Green axis of the CNRS GDR RSD (Network and Distributed Systems working group).
- Anne-Cécile Orgerie is secretary of the ASF: the French chapter of ACM SIGOPS.
- Cédric Tedeschi is a member of the steering committee of the Compas conference.

# 10.1.6. Scientific Expertise

Christine Morin was a member of the junior researcher selection committee at Inria Rennes Bretagne Atlantique.

Christine Morin was a member of the selection committee for an assistant-professor position at the University of Lyon.

Guillaume Pierre and Martin Quinson were members of the selection committee for a professor position at the University of Rennes 1.

Anne-Cécile Orgerie was the coordinator of the scientific challenge on "Digitizing Energy" for the Inria's scientific strategic plan 2018-2022.

Anne-Cécile Orgerie was a member of the selection committee for two assistant professor positions at IMT-Atlantique in Nantes.

Anne-Cécile Orgerie was a member of the admission jury for the *second concours* to recruit normalien students at Ecole Normale Supérieure of Rennes.

Jean-Louis Pazat is the coordinator of experts in Information Technology for the evaluation of international bilateral collaborations at the ministry of research and education.
Nikos Parlavantzas acted as an expert reviewer for ANR and SNSF projects.

#### 10.1.7. Research Administration

- Christine Morin is a member of the board of the Project-Team Committee of Inria Rennes Bretagne Atlantique.
- Christine Morin is a member of the University of Rennes 1 board of directors and of the International Affairs Commission and its board.
- Anne-Cécile Orgerie is officer (chargée de mission) for the IRISA cross-cutting axis on Green IT.
- Martin Quinson is the leader of the "Large Scale Systems" department of IRISA.
- Louis Rilling is a member of the scientific board of the research component ("club recherche") of "Pôle d'Excellence Cyber".

## 10.2. Teaching - Supervision - Juries

#### 10.2.1. Teaching

Christine Morin is responsible for the "Advanced Cloud Infrastructures" (ACI) teaching unit of the Master in Computer Science (SIF) of the University of Rennes 1 and of the EIT Digital Master School at the University of Rennes 1.

Christine Morin:

- Master 1: FreeRoom project (spanning the two semesters) co-supervised with Cédric Tedeschi, Adrien Capaine, and Paul Couderc, University of Rennes 1, France.
- Master 2: Advanced Computing Infrastructures" (ACI), 15 hours ETD, EIT Digital Master School, University of Rennes 1, France.

Anne-Cécile Orgerie:

- Master 1: Cloud & Big Data 12 hours of lecture and 12 hours of practical sessions at ENS Rennes
- Master 2: Green ICT 3 hours of invited lecture at Telecom SudParis
- Master 2: Green ICT 15 hours of lecture at IMT-Atlantique Nantes

Guillaume Pierre (at the University of Rennes 1):

- License 3 MIAGE: Systèmes (25 hours ETD)
- License 3 informatique: Organisation et utilisation des systèmes d'exploitation 2 (67 hours ETD)
- Master 1: Service Technologies (24 hours ETD)
- Master 1: Approche algorithmique des applications et systèmes répartis (32 hours ETD)
- Master 2: Techniques de développement logiciel dans le Cloud (39 hours ETD)
- Master 2: Advanced Cloud Infrastructures (10 hours ETD)

Martin Quinson (at ENS Rennes):

- Licence 3: Programming and Software Engineering (30 hours ETD); ARCSYS architecture et systèmes (60 hours ETD); Pedagogy (15 hours ETD).
- Agrégation Science Industrielle: Programming and Software Engineering (20 hours ETD); Operating Systems and C programming (20 hours ETD); Networking (20 hours ETD).
- Master 2: Pedagogy and Scientific Mediation for Computer Science (30 hours EDT)

Jean-Louis Pazat (at INSA Rennes):

- Licence 3: Parallel and Multicore Programming (36 hours ETD)
- Master 1: Parallel and Distributed Programming (36 hours ETD)
- Master 1: Scientific Parallel Programming (36 hours ETD)
- Licence 3: Networks and SOA (20 hours ETD)
- Master1 / Master 2: Introduction to IoT (24 hours ETD)
- Master2: Student Project in Cloud Computing (12 hours ETD)

Nikos Parlavantzas (at INSA Rennes):

- Master 2: Project in Large-Scale Systems (26 hours ETD)
- Master 1: Clouds (20 hours ETD)
- Master 1: Performance Evaluation (32 hours ETD)
- Master1/2: Introduction to IoT (8 hours ETD)
- Master 1: Operating Systems (36 hours ETD)
- Master 1: Big Data and Applications (15 hours ETD)
- Master 1: Parallel and Distributed Programming (12 hours ETD)
- Licence 3: Multicore Programming (12 hours ETD)
- Licence 3: Networks and SOA (14 hours ETD)
- Master 1: Software Development Project (30 hours ETD)

Cédric Tedeschi ( 220 hours ETD at Univ. Rennes 1):

- Master 1: Cooperation and concurrency in Systems and Networks
- Master 1: Parallel programming
- Master 2: Internet of Services and Infrastructures

#### E-learning

MOOC : Christine Morin, *Défis technologiques des villes intelligentes participatives*, 1 week in March 2017 and 1 week in November 2017, FUN (http://www.fun-mooc.fr/courses/inria/41009S02/session02/about), Inria MOOC-Lab, leaders, engineers, students, continuing education, respectively 3857 and 3022 registered participants for the March and November sessions.

SPOC : Christine Morin, Technological challenges of participatory smart cities, 1 week, eitdigitalx, EIT Digital professional school, leaders and engineers, continuing education, unknown number of registered participants

Pedagogical resources : Christine Morin, cloud-based urban data management, video and quizz, Master, https://www.eitdigital.eu/news-events/news/article/eit-digital-professional-school-launches-a-smart-cities-course-for-companies-and-cities/

#### 10.2.2. Supervision

HdR: Cédric Tedeschi, Distributed Chemically-Inspired Runtimes for Large Scale Adaptive Computing Platforms, defended on April 11th, 2017.

PhD: Anna Giannakou, Self-adaptable security monitoring for IaaS Cloud environments, defended on July 6th, 2017, Louis Rilling, Christine Morin, Jean-Louis Pazat.

PhD: Yunbo Li, Resource allocation in a Cloud partially powered by renewable energy sources, defended on June 12th, 2017, Anne-Cécile Orgerie, Jean-Marc Menaud (Ascola).

PhD: Sabbir Hasan, SLA Driven Cloud autoscaling for optimizing energy footprint, defended on May 3rd, 2017, Thomas Ledoux (Ascola), Jean-Louis Pazat.

PhD: Ismael Cuadrado Cordero, Energy-efficient and network-aware resource allocation in Cloud infrastructures, defended on February 9th, 2017, Christine Morin, Anne-Cécile Orgerie.

PhD in progress: Mehdi Belkhiria, Dynamic Stream Processing for Maritime Traffic Surveillance, started in December 2017, advised by Cédric Tedeschi and Christine Morin

PhD in progress: Ali Jawad Fahs, "Decentralized Fog Computing Infrastructure Control", started in October 2017, Guillaume Pierre.

PhD in progress: Clément El Baz, Reactive security monitoring in clouds, started in October 2017, Louis Rilling, Christine Morin.

PhD in progress: Loic Guegan, Simulating Internet of Things, started in October 2017, advised by Martin Quinson and Anne-Cécile Orgerie

PhD in progress (*co-tutelle*): Yasmina Bouizem, Energy-efficient, fault-tolerance mechanisms for containerized cloud applications, started in November 2016, Didi Fedoua (Tlemcen University, Algeria), Djawida Dib (Tlemcen University, Algeria), Christine Morin, Nikos Parlavantzas.

PhD in progress: The Anh Pham, Dynamic Formal Verification of High Performance Runtimes and Applications, started in November 2016, Martin Quinson, Thierry Jéron.

PhD in progress: Arif Ahmed, "Scalable Decentralized Edge Cloud Infrastructures", started in October 2016, Guillaume Pierre.

PhD in progress: Baptiste Goupille-Lescar, Designing agile, distributed cyber-physical systems with advanced collaboration capabilities, started in January 2016, Eric Lenormand (Thales), Christine Morin, Nikos Parlavantzas.

PhD in progress: Pernelle Mensah, Security policy adaptation driven by risk evaluation in modern communication infrastructures, started in December 2015, Samuel Dubus (Alcatel-Lucent), Christine Morin, Guillaume Piolle (Cidre), Eric Totel (Cidre).

PhD in progress: Genc Tato, Locality-aware Lazy Overlay Networks for WANS, started in December 2015, Marin Bertier, Cédric Tedeschi.

PhD in progress: Amir Teshome, Definition and enforcement of Service-Level Agreements for Cloud security monitoring, started in October 2015, Louis Rilling, Christine Morin.

PhD in progress: Issam Rais, Multi criteria scheduling for large scale High Performance Computing environments, started in October 2015, Anne-Cécile Orgerie, Anne Benoit (ROMA), Laurent Lefèvre (Avalon).

PhD in progress: David Guyon, Supporting energy-awareness for cloud users, started in September 2015, Anne-Cécile Orgerie, Christine Morin.

PhD in progress: Bruno Stevant, Resource allocation strategies for service distribution at the Internet edge to optimize end-to-end latency, started in December 2014 (part-time), Jean-Louis Pazat.

#### 10.2.3. Juries

- Christine Morin was a member in the HdR committee of Cédric Tedeschi, University of Rennes 1, April 11th, 2017.
- Christine Morin chaired the PhD committee of Sabbir Hasan, Insa Rennes, May 3rd, 2017.
- Christine Morin was an external reviewer in the PhD committee of Jiajun Cao, Northeastern University, September 2017.
- Christine Morin was the external reviewer in the PhD committee of Faiza Samreen, Lancaster University, November 23rd, 2017.
- Christine Morin chaired the PhD committee of Pedro Silva, ENS Lyon, December 11th, 2017.
- Anne-Cécile Orgerie was a member in the PhD committee of Millian Poquet, Université de Grenoble, December 19th, 2017.
- Anne-Cécile Orgerie was a member in the PhD committee of Inès de Courchelle, Université de Toulouse 3, November 20th, 2017.
- Jean-Louis Pazat chaired the PhD committee of Jean-Baptiste Barreau, INSA de Rennes, July 10, 2017.
- Jean-Louis Pazat was an external reviewer in the PhD committee of Riadh Karchoud, December 14, 2017.
- Guillaume Pierre was a member in the PhD committee of George Ioannidis, Ecole Polytechnique Fédérale de Lausanne, December 4th, 2017.

#### **10.3.** Popularization

Martin Quinson is a member of the ISO national working group (Info Sans Ordi – CS without computer), that aims at inventing, improving and spread unplug activities. These outreach activities are meant to introduce fundamental concepts of Computer Science in schools without the need of any electrical devices. In 2017, the group produced a special issue to the Tangente magazine (read by maths teachers in France). Martin Quinson authored two articles in this special issue.

Martin Quinson is on the scientific board of the Blaise Pascal foundation, boosting the scientific outreach in the domains of Maths and Computer Science. He participated this year to a working group aiming at the development of an online game that teaches maths to pupils.

Martin Quinson was in the scientific organizing committees for several outreach events toward the popularization of Maths and Computer Science this year: «Maths Vivantes» had a very general audience: a booth on a public space of Rennes for one afternoon (March). «Maths C2+» invited about 20 pupils invited at ENS Rennes for 3 days (March). «Metier en tout Genre» invited several dozen of high-school students for one day on the Rennes university campus (November).

# 11. Bibliography

#### Major publications by the team in recent years

- [1] S. COSTACHE, S. KORTAS, C. MORIN, N. PARLAVANTZAS.Market-based Autonomous Resource and Application Management in Private Clouds, in "Journal of Parallel and Distributed Computing", February 2017, vol. 100, p. 85-102 [DOI : 10.1016/J.JPDC.2016.10.003], https://hal.archives-ouvertes.fr/hal-01378536.
- [2] A. DEGOMME, A. LEGRAND, G. MARKOMANOLIS, M. QUINSON, M. L. STILLWELL, F. SUTER. *Simulating MPI applications: the SMPI approach*, in "IEEE Transactions on Parallel and Distributed Systems", February 2017, 14 [DOI: 10.1109/TPDS.2017.2669305], https://hal.inria.fr/hal-01415484.
- [3] E. FELLER, L. RAMAKRISHNAN, C. MORIN. Performance and energy efficiency of big data applications in cloud environments: A Hadoop case study, in "Journal of Parallel and Distributed Computing", 2015, vol. 79-80, p. 80-89 [DOI: 10.1016/J.JPDC.2015.01.001], https://hal.archives-ouvertes.fr/hal-01271141.
- [4] E. FELLER, L. RILLING, C. MORIN. Energy-Aware Ant Colony Based Workload Placement in Clouds, in "The 12th IEEE/ACM International Conference on Grid Computing (GRID-2011)", Lyon, France, September 2011, https://hal.inria.fr/inria-00626042.
- [5] A. IORDACHE, C. MORIN, N. PARLAVANTZAS, P. RITEAU. Resilin: Elastic MapReduce over Multiple Clouds, Inria, October 2012, n<sup>o</sup> RR-8081, https://hal.inria.fr/hal-00737030.
- [6] A. IORDACHE, G. PIERRE, P. SANDERS, J. G. D. F. COUTINHO, M. STILLWELL.*High Performance in the Cloud with FPGA Groups*, in "9th IEEE/ACM International Conference on Utility and Cloud Computing (UCC 2016)", Shanghai, China, December 2016, https://hal.inria.fr/hal-01356998.
- [7] A.-C. ORGERIE, M. DIAS DE ASSUNCAO, L. LEFÈVRE. A Survey on Techniques for Improving the Energy Efficiency of Large Scale Distributed Systems, in "ACM Computing Surveys", December 2014, vol. 46, n<sup>o</sup> 4, To appear, https://hal.inria.fr/hal-00767582.

[8] J. ROJAS BALDERRAMA, M. SIMONIN, C. TEDESCHI. GinFlow: A Decentralised Adaptive Workflow Execution Manager, in "30th IEEE International Parallel & Distributed Processing Symposium", Chicago, United States, Proceedings of the 30th IEEE International Parallel & Distributed Processing Symposium, May 2016, https://hal.archives-ouvertes.fr/hal-01292135.

#### **Publications of the year**

#### **Doctoral Dissertations and Habilitation Theses**

- [9] I. CUADRADO-CORDERO.*Microclouds : an approach for a network-aware energy-efficient decentralised cloud*, Université Rennes 1, February 2017, https://tel.archives-ouvertes.fr/tel-01498190.
- [10] A. GIANNAKOU. Self-adaptable Security Monitoring for IaaS Cloud Environments, Insa Rennes, July 2017, https://hal.inria.fr/tel-01653831.
- [11] M. S. HASAN. Smart management of renewable energy in clouds : from infrastructure to application, INSA de Rennes, May 2017, https://tel.archives-ouvertes.fr/tel-01617963.
- [12] Y. LI.*Resource allocation in a Cloud partially powered by renewable energy sources*, Ecole nationale supérieure Mines-Télécom Atlantique, June 2017, https://tel.archives-ouvertes.fr/tel-01595953.
- [13] C. TEDESCHI. Distributed Chemically-Inspired Runtimes for Large Scale Adaptive Computing Platforms, Université de Rennes 1, April 2017, Habilitation à diriger des recherches, https://tel.archives-ouvertes.fr/tel-01665776.

#### **Articles in International Peer-Reviewed Journal**

- [14] F. ALMEIDA, M. DIAS DE ASSUNCAO, J. BARBOSA, V. BLANCO, I. BRANDIC, G. DA COSTA, M. F. DOLZ, A. C. ELSTER, M. JARUS, H. KARATZA, L. LEFÈVRE, I. MAVRIDIS, A. OLEKSIAK, A.-C. ORGERIE, J.-M. PIERSON. *Energy Monitoring as an Essential Building Block Towards Sustainable Ultrascale Systems*, in "Sustainable Computing : Informatics and Systems", 2017, https://hal.inria.fr/hal-01627757.
- [15] N. BELDICEANU, B. DUMAS FERIS, P. GRAVEY, M. S. HASAN, C. JARD, T. LEDOUX, Y. LI, D. LIME, G. MADI-WAMBA, J.-M. MENAUD, P. MOREL, M. MORVAN, M.-L. MOULINARD, A.-C. ORGERIE, J.-L. PAZAT, O. H. ROUX, A. SHARAIHA.*Towards energy-proportional Clouds partially powered by renewable energy*, in "Computing", January 2017, vol. 99, n<sup>o</sup> 1, 20 [DOI: 10.1007/s00607-016-0503-z], https://hal. inria.fr/hal-01340318.
- [16] A. BENOIT, L. LEFÈVRE, A.-C. ORGERIE, I. RAÏS.Reducing the energy consumption of large scale computing systems through combined shutdown policies with multiple constraints, in "International Journal of High Performance Computing Applications", January 2018, vol. 32, n<sup>o</sup> 1, p. 176-188 [DOI: 10.1177/1094342017714530], https://hal.inria.fr/hal-01557025.
- [17] M. CALLAU-ZORI, L. SAMOILA, A.-C. ORGERIE, G. PIERRE. An experiment-driven energy consumption model for virtual machine management systems, in "Sustainable Computing : Informatics and Systems", 2017, https://hal.archives-ouvertes.fr/hal-01632962.
- [18] S. COSTACHE, D. DIB, N. PARLAVANTZAS, C. MORIN. Resource Management in Cloud Platform as a Service Systems: Analysis and Opportunities, in "Journal of Systems and Software", May 2017 [DOI: 10.1016/J.JSS.2017.05.035], https://hal.archives-ouvertes.fr/hal-01534992.

- [19] S. COSTACHE, S. KORTAS, C. MORIN, N. PARLAVANTZAS.*Market-based Autonomous Resource and Application Management in Private Clouds*, in "Journal of Parallel and Distributed Computing", February 2017, vol. 100, p. 85-102 [DOI : 10.1016/J.JPDC.2016.10.003], https://hal.archives-ouvertes.fr/hal-01378536.
- [20] A. DEGOMME, A. LEGRAND, G. MARKOMANOLIS, M. QUINSON, M. L. STILLWELL, F. SUTER. *Simulating MPI applications: the SMPI approach*, in "IEEE Transactions on Parallel and Distributed Systems", February 2017, 14 [DOI: 10.1109/TPDS.2017.2669305], https://hal.inria.fr/hal-01415484.
- [21] D. DIB, N. PARLAVANTZAS, C. MORIN.SLA-based PaaS profit optimization, in "Concurrency and Computation: Practice and Experience", August 2017 [DOI : 10.1002/CPE.4251], https://hal.archives-ouvertes.fr/ hal-01585288.
- [22] M. S. HASAN, F. ALVARES, T. LEDOUX, J.-L. PAZAT. Investigating Energy Consumption and Performance Trade-off for Interactive Cloud Application, in "IEEE Transactions on Sustainable Computing", June 2017 [DOI: 10.1109/TSUSC.2017.2714959], https://hal.archives-ouvertes.fr/hal-01540159.
- [23] M. S. HASAN, Y. KOUKI, T. LEDOUX, J.-L. PAZAT. Exploiting Renewable sources: when Green SLA becomes a possible reality in Cloud computing, in "IEEE Transactions on Cloud Computing", June 2017, vol. 5, n<sup>o</sup> 2, p. 249-262 [DOI: 10.1109/TCC.2015.2459710], https://hal.archives-ouvertes.fr/hal-01187907.
- [24] M. KURPICZ, A.-C. ORGERIE, A. SOBE, P. FELBER. Energy-proportional Profiling and Accounting in Heterogeneous Virtualized Environments, in "Sustainable Computing : Informatics and Systems", 2017, p. 1-36, https://hal.archives-ouvertes.fr/hal-01633435.
- [25] Y. LI, A.-C. ORGERIE, I. RODERO, B. LEMMA AMERSHO, M. PARASHAR, J.-M. MENAUD.End-toend Energy Models for Edge Cloud-based IoT Platforms: Application to Data Stream Analysis in IoT, in "Future Generation Computer Systems", 2017, p. 1-12 [DOI: 10.1016/J.FUTURE.2017.12.048], https:// hal.archives-ouvertes.fr/hal-01673501.

#### **International Conferences with Proceedings**

- [26] A. BENOIT, L. LEFÈVRE, A.-C. ORGERIE, I. RAÏS.Shutdown Policies with Power Capping for Large Scale Computing Systems, in "Euro-Par: International European Conference on Parallel and Distributed Computing", Santiago de Compostela, Spain, F. F. RIVERA, T. F. PENA, J. C. CA-BALEIRO (editors), Lecture Notes in Computer Science, springer, August 2017, vol. 10417, p. 134 -146 [DOI: 10.1109/COMST.2016.2545109], https://hal.archives-ouvertes.fr/hal-01589555.
- [27] B. CAMUS, F. DUFOSSÉ, A.-C. ORGERIE. A stochastic approach for optimizing green energy consumption in distributed clouds, in "SMARTGREENS 2017 - International Conference on Smart Cities and Green ICT Systems", Porto, Portugal, April 2017, https://hal.inria.fr/hal-01475431.
- [28] R.-A. CHERRUEAU, D. PERTIN, A. SIMONET, A. LEBRE, M. SIMONIN. Toward a Holistic Framework for Conducting Scientific Evaluations of OpenStack, in "The 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)", Madrid, Spain, May 2017, p. 544-548 [DOI: 10.1109/CCGRID.2017.87], https://hal.archives-ouvertes.fr/hal-01664515.
- [29] I. CUADRADO-CORDERO, A.-C. ORGERIE, J.-M. MENAUD.Comparative Experimental Analysis of the Quality-of-Service and Energy-Efficiency of VMs and Containers' Consolidation for Cloud Applications,

in "SoftCOM: International Conference on Software, Telecommunications and Computer Networks", Split, Croatia, September 2017, https://hal.archives-ouvertes.fr/hal-01578325.

- [30] L. GUEGAN, N. MURAD, J. M. LEBRETON, S. BONHOMMEAU. Integrating Over Sea Radio Channel for Sea Turtles Localization in the Indian Ocean, in "IEEE RADIO 2017", Cape Town, South Africa, September 2017, https://arxiv.org/abs/1712.03146, https://hal.archives-ouvertes.fr/hal-01632295.
- [31] D. GUYON, A.-C. ORGERIE, C. MORIN, D. AGARWAL. How Much Energy can Green HPC Cloud Users Save?, in "PDP 2017 - 25th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing", Saint Petersbourg, Russia, March 2017, https://hal.inria.fr/hal-01439874.
- [32] D. GUYON, A.-C. ORGERIE, C. MORIN.GLENDA: Green Label towards Energy proportioNality for IaaS DAta centers, in "E2DC: International Workshop on Energy Efficient Data Centres (e-Energy Workshop)", Hong Kong, Hong Kong SAR China, May 2017 [DOI: 10.1145/3077839.3084028], https://hal.archivesouvertes.fr/hal-01514948.
- [33] M. S. HASAN, F. ALVARES DE OLIVEIRA JR., T. LEDOUX. GPaaScaler: Green Energy aware Platform Scaler for Interactive Cloud Application, in "IEEE/ACM International Conference on Utility and Cloud Computing", Austin, United States, December 2017, https://hal.archives-ouvertes.fr/hal-01650136.
- [34] Best Paper

T. HAUDEBOURG, A.-C. ORGERIE. On the Energy Efficiency of Sleeping and Rate Adaptation for Network Devices, in "ICA3PP 2017 - 17th International Conference on Algorithms and Architectures for Parallel Processing", Helsinki, Finland, Springer, August 2017, vol. 10393, p. 132-146, Best Paper Award [DOI: 10.1007/978-3-319-65482-9\_9], https://hal.archives-ouvertes.fr/hal-01575513.

- [35] F. C. HEINRICH, T. CORNEBIZE, A. DEGOMME, A. LEGRAND, A. CARPEN-AMARIE, S. HUNOLD, A.-C. ORGERIE, M. QUINSON. Predicting the Energy Consumption of MPI Applications at Scale Using a Single Node, in "Cluster 2017", Hawaii, United States, IEEE, September 2017, https://hal.inria.fr/hal-01523608.
- [36] Y. LI, A.-C. ORGERIE, J.-M. MENAUD. Balancing the use of batteries and opportunistic scheduling policies for maximizing renewable energy consumption in a Cloud data center, in "PDP 2017 - 25th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing", St Petersburg, Russia, March 2017, https://hal.inria.fr/hal-01432752.
- [37] Y. LI, A.-C. ORGERIE, I. RODERO, M. PARASHAR, J.-M. MENAUD. Leveraging Renewable Energy in Edge Clouds for Data Stream Analysis in IoT, in "CCGrid 2017 - IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing", Madrid, Spain, May 2017, https://hal.inria.fr/hal-01472358.
- [38] G. MADI WAMBA, Y. LI, A.-C. ORGERIE, N. BELDICEANU, J.-M. MENAUD. Cloud workload prediction and generation models, in "SBAC-PAD: International Symposium on Computer Architecture and High Performance Computing", Campinas, Brazil, October 2017, https://hal.archives-ouvertes.fr/hal-01578354.
- [39] G. MADI WAMBA, Y. LI, A.-C. ORGERIE, N. BELDICEANU, J.-M. MENAUD.Green energy aware scheduling problem in virtualized datacenters, in "ICPADS 2017", Shenzen, China, December 2017, https:// hal.archives-ouvertes.fr/hal-01582936.

- [40] D. MARGERY, D. GUYON, A.-C. ORGERIE, C. MORIN, G. FRANCIS, C. PALANSURIYA, K. KAVOUS-SANAKIS.A CO2 emissions accounting framework with market-based incentives for Cloud infrastructures, in "SMARTGREENS: International Conference on Smart Cities and Green ICT Systems", Porto, Portugal, April 2017, https://hal.inria.fr/hal-01486185.
- [41] P. MENSAH, S. DUBUS, W. KANOUN, C. MORIN, G. PIOLLE, E. TOTEL. Connectivity Extraction in Cloud Infrastructures, in "CNSM 2017 - 13th International Conference on Network and Service Management", Tokyo, Japan, November 2017, p. 1-5, https://hal.inria.fr/hal-01593346.
- [42] P. MENSAH, S. DUBUS, W. KANOUN, C. MORIN, G. PIOLLE, E. TOTEL. Connectivity Graph Reconstruction for Networking Cloud Infrastructures, in "NCA 2017 - 16th IEEE International Symposium on Network Computing and Applications", Cambridge, MA, United States, October 2017, https://hal.inria.fr/hal-01612988.
- [43] A.-C. ORGERIE, B. LEMMA AMERSHO, T. HAUDEBOURG, M. QUINSON, M. RIFAI, D. LOPEZ PACHECO, L. LEFÈVRE. Simulation Toolbox for Studying Energy Consumption in Wired Networks, in "CNSM: International Conference on Network and Service Management", Tokyo, Japan, November 2017, p. 1-5, https://hal. archives-ouvertes.fr/hal-01630226.
- [44] T. A. PHAM, T. JÉRON, M. QUINSON. Verifying MPI Applications with SimGridMC, in "Correctness 2017 First International Workshop on Software Correctness for HPC Applications", Denver, United States, November 2017 [DOI: 10.1145/3145344.3145345], https://hal.inria.fr/hal-01632421.
- [45] M. L. PHAM, N. PARLAVANTZAS, C. MORIN, S. ARNOUX, L. QI, P. GONTIER, P. EZANNO.DiFFuSE, a Distributed Framework for Cloud-based Epidemic Simulations: a Case Study in Modelling the Spread of Bovine Viral Diarrhea Virus, in "CloudCom 2017 - 9th IEEE International Conference on Cloud Computing Technology and Science", Hong-Kong, China, December 2017, https://hal.inria.fr/hal-01612979.
- [46] C. RUIZ, H. A. DURAN-LIMON, N. PARLAVANTZAS. An RLS Memory-based Mechanism for the Automatic Adaptation of VMs on Cloud Environments, in "Workshop on Adaptive Resource Management and Scheduling for Cloud Computing", Washington, DC, United States, July 2017 [DOI : 10.1145/3110355.3110358], https://hal.archives-ouvertes.fr/hal-01637794.
- [47] G. TATO, M. BERTIER, C. TEDESCHI. Designing Overlay Networks for Decentralized Clouds, in "Future-Cloud 2017 - International Workshop on the Future of Cloud Computing and Cloud Services", Hong-Kong, China, December 2017, https://hal.inria.fr/hal-01634409.
- [48] A. VAN KEMPEN, T. CRIVAT, B. TRUBERT, D. ROY, G. PIERRE.*MEC-ConPaaS: An experimental single-board based mobile edge cloud*, in "IEEE Mobile Cloud Conference", San Francisco, United States, April 2017, https://hal.inria.fr/hal-01446483.
- [49] A. T. WONJIGA, L. RILLING, C. MORIN. Verification for Security Monitoring SLAs in IaaS Clouds: the Example of a Network IDS, in "NOMS 2018 - Network Operations and Management Symposium", Taipei, Taiwan, April 2018, https://hal.inria.fr/hal-01663941.

#### **National Conferences with Proceeding**

[50] S. DELAMARE, P. MORILLON, L. NUSSBAUM. Réalisation d'expériences avec Grid'5000, in "JRES2017 -Journées Réseaux de l'enseignement et de la recherche", Nantes, France, November 2017, https://hal.inria.fr/ hal-01639524.

#### **Conferences without Proceedings**

[51] N. MURAD, L. GUEGAN, S. BONHOMMEAU. Why satellite localization beacons are not adapted for marine turtles' study: a sea wireless sensors network solution, in "GIIS'17 - Global Information Infrastructure and Networking Symposium", Saint Pierre, Réunion, IEEE, October 2017, https://hal.archives-ouvertes.fr/hal-01631305.

#### Scientific Books (or Scientific Book chapters)

[52] J. G. F. COUTINHO, M. STILLWELL, K. ARGYRAKI, G. IOANNIDIS, A. IORDACHE, C. KLEINEWEBER, A. KOLIOUSIS, J. MCGLONE, G. PIERRE, C. RAGUSA, P. SANDERS, T. SCHÜTT, T. YU, A. WOLF.*The HARNESS Platform: A Hardware-and Network-Enhanced Software System for Cloud Computing*, in "Software Architecture for Big Data and the Cloud", I. MISTRIK, R. BAHSOON, N. ALI, M. HEISEL, B. MAXIM (editors), Morgan Kaufmann, June 2017, https://hal.inria.fr/hal-01507344.

#### **Research Reports**

- [53] R.-A. CHERRUEAU, M. SIMONIN, A. VAN KEMPEN.*EnosStack: A LAMP-like stack for the experimenter*, Inria Rennes - Bretagne Atlantique, January 2018, n<sup>o</sup> RR-9146, p. 1-14, https://hal.inria.fr/hal-01689726.
- [54] L. M. PHAM, N. PARLAVANTZAS, C. MORIN, S. ARNOUX, L. QI, P. GONTIER, P. EZANNO.DiFFuSE, a Distributed Framework for Cloud-based Epidemic Simulations: a Case Study in Modelling the Spread of Bovine Viral Diarrhea Virus, Inria Rennes - Bretagne Atlantique, September 2017, n<sup>o</sup> RR-9094, https://hal. archives-ouvertes.fr/hal-01586945.
- [55] A. TESHOME, L. RILLING, C. MORIN. Verification for Security Monitoring SLAs in IaaS Clouds: the Example of a Network IDS, Inria Rennes Bretagne Atlantique, August 2017, n<sup>o</sup> RR-9091, https://hal.inria.fr/hal-01577814.

#### **Other Publications**

- [56] R. FABLET, N. BELLEC, L. CHAPEL, C. FRIGUET, R. GARELLO, P. GLOAGUEN, G. HAJDUCH, S. LEFÈVRE, F. MERCIOL, P. MORILLON, C. MORIN, M. SIMONIN, R. TAVENARD, C. TEDESCHI, R. VADAINE.Next Step for Big Data Infrastructure and Analytics for the Surveillance of the Maritime Traffic from AIS & Sentinel Satellite Data Streams, November 2017, p. 1-4, BiDS' 2017 Conference on Big Data from Space, Poster, https://hal.inria.fr/hal-01671323.
- [57] B. GOUPILLE-LESCAR, E. LENORMAND, C. MORIN, N. PARLAVANTZAS. Introducing Dynamic Adaptation in High Performance Real-Time Computing Platforms for Sensors, September 2017, ANDARE 2017 - 1st Workshop on AutotuniNg and aDaptivity AppRoaches for Energy efficient HPC Systems, Poster, https://hal. inria.fr/hal-01624262.
- [58] F. C. HEINRICH, A. CARPEN-AMARIE, A. DEGOMME, S. HUNOLD, A. LEGRAND, A.-C. ORGERIE, M. QUINSON. Predicting the Performance and the Power Consumption of MPI Applications With SimGrid, January 2017, working paper or preprint, https://hal.inria.fr/hal-01446134.

#### **References in notes**

[59] Automatiser la reconfiguration des sondes de détection d'intrusion pour le cloud, July 2017, n<sup>o</sup> 47, http:// emergences.inria.fr/2017/newsletter-n47/newsletter\_show\_all. [60] A. LÈBRE, J. PASTOR, M. BERTIER, F. DESPREZ, J. ROUZAUD-CORNABAS, C. TEDESCHI, P. ANEDDA, G. ZANETTI, R. NOU, T. CORTES, E. RIVIÈRE, T. ROPARS. Beyond The Cloud, How Should Next Generation Utility Computing Infrastructures Be Designed?, Inria, July 2013, n<sup>o</sup> RR-8348, https://hal.inria.fr/ hal-00854204.

# **Project-Team PACAP**

# Pushing Architecture and Compilation for Application Performance

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: Université Rennes 1

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Architecture, Languages and Compilation

# **Table of contents**

1.	Personnel	737	
2.	Overall Objectives	738	
	2.1.1. Long-Term Goal	738	
	2.1.2. Approach	739	
	2.1.3. Latency-oriented Computing	739	
	2.1.4. Throughput-Oriented Computing	739	
	2.1.5. Real-Time Systems – WCET	739	
	2.1.6. Performance Assessment	740	
	2.1.7. Dealing with Faults – Reliability	740	
	2.1.8. Dealing with Attacks – Security	741	
	2.1.9. Green Computing	741	
3.	Research Program	741	
	3.1. Motivation	741	
	3.1.1. Technological constraints	741	
	3.1.2. Evolving community	742	
	3.1.3. Domain constraints	742	
	3.2. Research Objectives	742	
	3.2.1. Static Compilation	743	
	3.2.2. Software Adaptation	743	
	3.2.3. Research directions in uniprocessor microarchitecture	744	
	3.2.4. Towards heterogeneous single-ISA CPU-GPU architectures	745	
	3.2.5. Real-time systems	746	
	3.2.6. Fault Tolerance	746	
	3.2.7. Power efficiency	747	
	3.2.8. Security	748	
4.	Application Domains	749	
5.	New Software and Platforms	749	
	5.1. ATMI	749	
	5.2. HEPTANE	749	
	5.3. tiptop	749	
	5.4. PADRONE	750	
5.5. If-memo		751	
	5.6. Simty	751	
	5.7. Barra	751	
6.	New Results	752	
	6.1. Compiler, vectorization, interpretation	752	
	6.1.1. Improving sequential performance through memoization	752	
	6.1.2. Optimization in the Presence of NVRAM	752	
	6.1.3. Dynamic Binary Optimization	753	
	6.1.3.1. Dynamic Function Specialization	753	
	6.1.3.2. Runtime Vectorization of Binary Programs	753	
	6.1.4. Hardware/Software JIT Compiler	753	
	6.1.5. Customized Precision Computing	754	
	6.1.6. SPMD Function Call Re-Vectorization		
	6.1.7. Qubit allocation for quantum circuit compilers	755	
	6.2. Processor Architecture	755	
	6.2.1.1. Bayesian TAGE predictors	755	
	6.2.1.2. Interactions Between Value Prediction and Compiler Optimizations	756	
	6.2.1.3. Prefetch Management on Multicore Systems	756	

	6.2.1.4. Managing Shared Last Level Caches in Large Multicores	756
	6.2.1.5. Augmenting superscalar architecture for efficient many-thread parallel executi	on 756
	6.2.1.6. Generalizing the SIMT execution model to general-purpose instruction sets	757
	6.2.1.7. Toward out-of-order SIMT microarchitecture	757
	6.3. WCET estimation and optimization	757
	6.3.1. WCET estimation for many core processors	757
	6.3.1.1. Optimization of WCETs by considering the effects of local caches	758
	6.3.1.2. Accounting for shared resource contentions to minimize WCETs	758
	6.3.1.3. WCET-Aware Parallelization of Model-Based Applications for Multi-Cores	758
	6.3.2. WCET estimation tool and benchmarks	759
	6.4. Security	759
7.	Bilateral Contracts and Grants with Industry	759
	7.1. Bilateral Contracts with Industry	759
	7.2. Bilateral Grants with Industry	760
	7.2.1. Intel research grant INTEL2014-8957	760
	7.2.2. Intel research grant INTEL2016-11174	760
8.	Partnerships and Cooperations	760
	8.1. Regional Initiatives	760
	8.2. National Initiatives	760
	8.2.1. Capacités: Projet "Investissement d'Avenir", 1/11/14 to 31/01/2018	760
	8.2.2. Zero Power Computing Systems (ZEP): Inria Project Lab, 2017–2020	761
	8.2.3. ANR Continuum 2015–2019	761
	8.2.4. ANR W-SEPT 2012-2017	761
	8.3. European Initiatives	761
	8.3.1.1. ANTAREX	761
	8.3.1.2. Eurolab-4-HPC	762
	8.3.1.3. ARGO	763
	8.3.1.4. HiPEAC4 NoE	763
	8.4. International Initiatives	764
	8.4.1. ANR CHIST-ERA SECODE 2016-2018	764
	8.4.2. PHC IMHOTEP	764
	8.4.3. Inria Associate Teams Not Involved in an Inria International Labs	764
	8.5. International Research Visitors	765
9.	Dissemination	
	9.1. Promoting Scientific Activities	765
	9.1.1. Scientific Events Organisation	765
	9.1.2. Scientific Events Selection	765
	9.1.2.1. Chair of Conference Program Committees	765
	9.1.2.2. Member of the Conference Program Committees	765
	9.1.2.3. Reviewer	/65
	9.1.3. Journal	/65
	9.1.3.1. Member of the Editorial Boards	/65
	9.1.3.2. Reviewer - Reviewing Activities	765
	9.1.4. Invited Talks	765
	9.1.5. Leadership within the Scientific Community	766
	9.1.6. Research Administration	766
	9.2. Teaching - Supervision - Juries	766
	9.2.1. Teaching	766
	9.2.2. Supervision	766
	9.2.3. Juries	/6/
	9.3. Popularization	/68

10.	Bibliography		768
-----	--------------	--	-----

### **Project-Team PACAP**

Creation of the Project-Team: 2016 July 01

#### **Keywords:**

#### **Computer Science and Digital Science:**

- A1.1. Architectures
- A1.1.1. Multicore, Manycore
- A1.1.2. Hardware accelerators (GPGPU, FPGA, etc.)
- A1.1.3. Memory models
- A1.1.4. High performance computing
- A1.1.5. Exascale
- A1.1.9. Fault tolerant systems
- A1.1.10. Reconfigurable architectures
- A1.1.11. Quantum architectures
- A1.6. Green Computing
- A2.2. Compilation
- A2.2.1. Static analysis
- A2.2.2. Memory models
- A2.2.3. Run-time systems
- A2.2.4. Parallel architectures
- A2.2.5. GPGPU, FPGA, etc.
- A2.2.6. Adaptive compilation
- A2.3.1. Embedded systems
- A2.3.3. Real-time systems
- A4.2. Correcting codes
- A4.4. Security of equipment and software
- A8.9. Performance evaluation
- A8.10. Computer arithmetic

#### **Other Research Topics and Application Domains:**

- B1. Life sciences
- B2. Health
- B3. Environment and planet
- B4. Energy
- B5. Industry of the future
- B6. IT and telecom
- B7. Transport and logistics
- **B8.** Smart Cities and Territories
- B9. Society and Knowledge

# 1. Personnel

**Research Scientists** 

Erven Rohou [Team leader, Inria, Senior Researcher, HDR] Sylvain Collange [Inria, Researcher] Pierre Michaud [Inria, Researcher] André Seznec [Inria, Senior Researcher, HDR]

#### **Faculty Members**

Damien Hardy [Univ de Rennes I, Associate Professor] Isabelle Puaut [Univ de Rennes I, Professor, HDR]

#### **Technical Staff**

Loïc Besnard [CNRS SED, Senior Research Engineer, seconded at 40 %] Nicolas Kiss [Inria] Imane Lasri [Inria] Sébastien Martinez [Univ de Rennes I, until Jun 2017] Kévin Le Bon [Inria, from Sep 2017]

#### **PhD Students**

Arif Ali Ana-Pparakkal [Inria] Rabab Bouziane [Inria] Niloofar Charmchi [Inria] Kleovoulos Kalaitzidis [Inria] Viet Anh Nguyen [Univ de Rennes I] Daniel Rodrigues Carvalho [Inria, from Oct 2017] Benjamin Rouxel [Univ de Rennes I]

#### **Post-Doctoral Fellows**

Fernando Endo [Inria, until May 2017] Imen Fassi [Univ de Rennes I, from Oct 2017] Byron Hawkins [Inria, from Sep 2017] Biswabandan Panda [Inria, until May 2017] Stefanos Skalistis [Univ de Rennes I, from Dec 2017]

#### Visiting Scientists

Stefano Cherubin [Politecnico di Milano, from Mar 2017 until Apr 2017] Andrei Rimsa Alvares [Centro Federal de Educação Tecnológica de Minas Gerais, until Feb 2017] Marcos Siraichi [Universidade Federal de Minas Gerais, from Dec 2017]

#### Administrative Assistant

Virginie Desroches [Inria]

# 2. Overall Objectives

#### 2.1. Overall Objectives

#### 2.1.1. Long-Term Goal

In brief, the long-term goal of the PACAP project-team is about *performance*, that is: how fast programs run. We intend to contribute to the ongoing race for exponentially increasing performance and for performance guarantees.

Traditionally, the term "performance" is understood as "how much time is needed to complete execution". *Latency*-oriented techniques focus on minimizing the average-case execution time (ACET). We are also interested in other definitions of performance. *Throughput*-oriented techniques are concerned with how many units of computations can be completed per unit of time. This is more relevant on manycores and GPUs where many computing nodes are available, and latency is less critical. Finally, we also study worst-case execution time (WCET), which is extremely important for critical real-time systems where designers must guarantee that deadlines are met, in any situation.

Given the complexity of current systems, simply assessing their performance has become a non-trivial task which we also plan to tackle.

We occasionally consider other metrics related to performance, such as power efficiency, total energy, overall complexity, and real-time response guarantee. Our ultimate goal is to propose solutions that make computing systems more efficient, taking into account current and envisioned applications, compilers, runtimes, operating systems, and microarchitectures. And since increased performance often comes at the expense of another metric, identifying the related trade-offs is of interest to PACAP.

The previous decade witnessed the end of the "magically" increasing clock frequency and the introduction of commodity multicore processors. PACAP will likely experience the end of Moore's law <sup>0</sup>, and the generalization of commodity heterogeneous manycore processors. This impacts how performance is increased and how it can be guaranteed. It is also a time where exogenous parameters should be promoted to first-class citizens:

- 1. the existence of faults, whose impact is becoming increasingly important when the photo-lithography feature size decreases;
- 2. the need for security at all levels of computing systems;
- 3. green computing, or the growing concern of power consumption.

#### 2.1.2. Approach

We strive to address performance in a way as transparent as possible for users. For example, instead of proposing any new language, we consider existing applications (written for example in standard C), and we develop compiler optimizations that immediately benefit programmers; we propose microarchitectural features as opposed to changes in processor instruction sets; we analyze and re-optimize binary programs automatically, without any user intervention.

The perimeter of research directions proposed for the PACAP project-team derive from the intersection of two axes: on the one hand, our high-level research objectives, derived from the overall panorama of computing systems, on the other hand the existing expertise and background of the team members on key technology (see illustration on Figure 1). Note that it does not imply that we will systematically explore all intersecting points of the figure, yet all correspond to a sensible research direction. These lists are neither exhaustive, nor final. Operating systems in particular constitute a promising operating point for several of the issues we plan to tackle. Other aspects will likely emerge during the lifespan of the project-team.

#### 2.1.3. Latency-oriented Computing

Improving the ACET of general purpose systems has been the "core business" of PACAP's ancestors (CAPS and ALF) for two decades. We plan to pursue this line of research, acting at all levels: compilation, dynamic optimizations, and microarchitecture.

#### 2.1.4. Throughput-Oriented Computing

The goal is to maximize the performance-to-power ratio. We will leverage the execution model of throughputoriented architectures (such as GPUs) and extend it towards general purpose systems. To address the memory wall issue, we will consider bandwidth saving techniques, such as cache and memory compression.

#### 2.1.5. Real-Time Systems – WCET

Designers of real-time systems must provide an upper bound of the worst-case execution time of the tasks within their systems. By definition this bound must be safe (i.e. greater than any possible execution time). To be useful, WCET estimates have to be as tight as possible. The process of obtaining a WCET bound consists in analyzing a binary executable, modeling the hardware, and then maximizing an objective function that takes into account all possible flows of execution and their respective execution times. Our research will consider the following directions:

<sup>&</sup>lt;sup>0</sup>Moore's law states that the number of transistors in a circuit doubles (approximately) every two years.



Figure 1. Perimeter of Research Objectives

- 1. better modeling of hardware to either improve tightness, or handle more complex hardware (e.g. multicores);
- 2. eliminate unfeasible paths from the analysis;
- 3. consider probabilistic approaches where WCET estimates are provided with a confidence level.

#### 2.1.6. Performance Assessment

Moore's law drives the complexity of processor micro-architectures, which impacts all other layers: hypervisors, operating systems, compilers and applications follow similar trends. While a small category of experts is able to comprehend (parts of) the behavior of the system, the vast majority of users are only exposed to – and interested in – the bottom line: how fast their applications are actually running. In the presence of virtual machines and cloud computing, multi-programmed workload add yet another degree of non-determinism to the measure of performance. We plan to research how application performance can be characterized and presented to a final user: behavior of the microarchitecture, relevant metrics, possibly visual rendering. Targeting our own community, we also research techniques appropriate for fast and accurate ways to simulate future architectures, including heterogeneous designs, such as latency/throughput platforms.

Once diagnosed, the way bottlenecks are addressed depends on the level of expertise of users. Experts can typically be left with a diagnostic as they probably know better how to fix the issue. Less knowledgeable users must be guided to a better solution. We plan to rely on iterative compilation to generate multiple versions of critical code regions, to be used in various runtime conditions. To avoid the code bloat resulting from multiversioning, we will leverage split-compilation to embed code generation "recipes" to be applied just-in-time, or even at rutime thanks to dynamic binary translation. Finally, we will explore the applicability of auto-tuning, where programmers expose which parameters of their code can be modified to generate alternate versions of the program (for example trading energy consumption for quality of service) and let a global orchestrator make decisions.

#### 2.1.7. Dealing with Faults – Reliability

Semiconductor technology evolution suggests that permanent failure rates will increase dramatically with scaling. While well-known approaches, such as error correcting codes, exist to recover from failures and provide fault-free chips, the exponential growth of the number of faults will make them unaffordable in the future. Consequently, other approaches like fine-grained disabling and reconfiguration of hardware elements (e.g. individual functional units or cache blocks) will become economically necessary. This fine-grained disabling will degrade performance compared to a fault-free execution. This evolution impacts performance (both ACET and WCET). We plan to address this evolution, and propose new techniques, which can be developed at any level. For example, at microarchitecture level, one might consider designing part of a cache in an older technology to guarantee a minimum level of performance; at compile-time, one might generate

redundant code for critical sections; at run-time, one can monitor faults and apply corrective measures to the software, or hardware. Solutions involving multiple levels are also very promising.

#### 2.1.8. Dealing with Attacks – Security

Computer systems are under constant attack, from young hackers trying to show their skills, to "professional" criminals stealing credit card information, and even government agencies with virtually unlimited resources. A vast amount of techniques have been proposed in the literature to circumvent attacks. Many of them cause significant slowdowns due to additional checks and countermeasures. Thanks to our expertise in microarchitecture and compilation techniques, we will be able to significantly improve efficiency, robustness and coverage of security mechanism, as well as to partner with field experts to design innovative solutions.

#### 2.1.9. Green Computing

Power consumption has become a major concern of computing systems, at all form factors, ranging from energy-scavenging sensors for IoT, to battery powered embedded systems and laptops, and up to supercomputers operating in the tens of megawatts. Execution time and energy are often related optimization goals. Optimizing for performance under a given power cap, however, introduces new challenges. It also turns out that technologists introduce new solutions (e.g. magnetic RAM) which, in turn, result in new trade-offs and optimization opportunities.

# 3. Research Program

#### 3.1. Motivation

Our research program is naturally driven by the evolution of our ecosystem. Relevant recent changes can be classified in the following categories: technological constraints, evolving community, and domain constraints. We hereby summarize these evolutions.

#### 3.1.1. Technological constraints

Until recently, binary compatibility guaranteed portability of programs, while increased clock frequency and improved micro-architecture provided increased performance. However, in the last decade, advances in technology and micro-architecture started translating into more parallelism instead. Technology roadmaps even predict the feasibility of thousands of cores on a chip by 2020. Hundreds are already commercially available. Since the vast majority of applications are still sequential, or contain significant sequential sections, such a trend put an end to the automatic performance improvement enjoyed by developers and users. Many research groups consequently focused on parallel architectures and compiling for parallelism.

Still, the performance of applications will ultimately be driven by the performance of the sequential part. Despite a number of advances (some of them contributed by members of the team), sequential tasks are still a major performance bottleneck. Addressing it is still on the agenda of the proposed PACAP project-team.

In addition, due to power constraints, only part of the billions of transistors of a microprocessor can be operated at any given time (the *dark silicon* paradigm). A sensible approach consists in specializing parts of the silicon area to provide dedicated accelerators (not run simultaneously). This results in diverse and heterogeneous processor cores. Application and compiler designers are now confronted with a moving target, challenging portability and jeopardizing performance.

Finally, we live in a world where billions of sensors, actuators, and computers play a crucial role in our life: flight control, nuclear plant management, defense systems, banking, or health care. These systems must be reliable, despite the fact that they are subject to faults (for example due to aging, charged particle hit, or random noise). Faults will soon become the new *de facto* standard. The evolutions of the semiconductor industry predict an exponential growth of the number of permanent faults [56]. Reliability considerations usually degrade performance. We will propose solutions to mitigate this impact (for example by limiting overheads to critical sections).

#### Note on technology.

Technology also progresses at a fast pace. We do not propose to pursue any research on technology *per se*. Recently proposed paradigms (non-Si, brain-inspired) have received lots of attention from the research community. We do *not* intend to invest in those paradigms, but we will continue to investigate compilation and architecture for more conventional programming paradigms. Still, several technological shifts may have consequences for us, and we will closely monitor their developments, they include for example non-volatile memory (impacts security, makes writes longer than loads), 3D-stacking (impacts bandwidth), and photonics (impacts latencies and connection network).

#### 3.1.2. Evolving community

The PACAP project-team tackles performance-related issues, for conventional programming paradigms. In fact, programming complex environments is no longer the exclusive domain of experts in compilation and architecture. A large community now develops applications for a wide range of targets, including mobile "apps", cloud, multicore or heterogeneous processors.

This also includes domain scientists (in biology, medicine, but also social sciences) who started relying heavily on computational resources, gathering huge amounts of data, and requiring considerable amount of processing to analyze them. Our research is motivated by the growing discrepancy between on the one hand the complexity of the workloads and the computing systems, and on the other hand the expanding community of developers at large, with limited expertise to optimize and to map efficiently computations to compute nodes.

#### 3.1.3. Domain constraints

Mobile, embedded systems have become ubiquitous. Many of them have real-time constraints. For this class of systems, correctness implies not only producing the correct result, but also doing so within specified deadlines. In the presence of heterogeneous, complex and highly dynamic systems, producing *tight* (i.e. useful) upper bound to the worst-case execution time has become extremely challenging. Our research will aim at improving the tightness as well as enlarging the set of features that can be safely analyzed.

The ever growing dependence of our economy on computing systems also implies that security has become of utmost importance. Many systems are under constant attacks from intruders. Protection has a cost also in terms of performance. We plan to leverage our background to contribute solutions that minimize this impact.

#### Note on Applications Domains.

PACAP works on fundamental technologies for computer science: processor architecture, performanceoriented compilation and guaranteed response time for real-time. The research results may have impacts on any application domain that requires high performance execution (telecommunication, multimedia, biology, health, engineering, environment...), but also on many embedded applications that exhibit other constraints such as power consumption, code size and guaranteed response time.

We strive to extract from active domains the fundamental characteristics that are relevant to our research. For example, *big data* is of interest to PACAP because it relates to the study of hardware/software mechanisms to efficiently transfer huge amounts of data to the computing nodes. Similarly, the *Internet of Things* is of interest because it has implications in terms of ultra low power consumption.

#### 3.2. Research Objectives

Processor micro-architecture and compilation have been at the core of the research carried by the members of the project teams for two decades, with undeniable contributions. They continue to be the foundation of PACAP.

Heterogeneity and diversity of processor architectures now require new techniques to guarantee that the hardware is satisfactorily exploited by the software. One of our goals is to devise new static compilation techniques (cf. Section 3.2.1), but also build upon iterative [1] and split [2] compilation to continuously adapt software to its environment (Section 3.2.2). Dynamic binary optimization will also play a key role in delivering adapting software and delivering performance.

The end of Moore's law and Dennard's scaling <sup>0</sup> offer an exciting window of opportunity, where performance improvements will no longer derive from additional transistor budget or increased clock frequency, but rather come from breakthroughs in microarchitecture (Section 3.2.3). Reconciling CPU and GPU designs (Section 3.2.4) is one of our objectives.

Heterogeneity and multicores are also major obstacles to determining tight worst-case execution times of real-time systems (Section 3.2.5), which we plan to tackle.

Finally, we also describe how we plan to address transversal aspects such as reliability (Section 3.2.6), power efficiency (Section 3.2.7), and security (Section 3.2.8).

#### 3.2.1. Static Compilation

Static compilation techniques continues to be relevant to address the characteristics of emerging hardware technologies, such as non-volatile memories, 3D-stacking, or novel communication technologies. These techniques expose new characteristics to the software layers. As an example, non-volatile memories typically have asymmetric read-write latencies (writes are much longer than reads) and different power consumption profiles. PACAP studies the new optimization opportunities and develop tailored compilation techniques for the upcoming compute nodes. New technologies may also be coupled with traditional solutions to offer new trade-offs. We study how programs can adequately exploit the specific features of the proposed heterogeneous compute nodes.

We propose to build upon iterative compilation [1] to explore how applications perform on different configurations. When possible, Pareto points are related to application characteristics. The best configuration, however, may actually depend on runtime information, such as input data, dynamic events, or properties that are available only at runtime. Unfortunately a runtime system has little time and means to determine the best configuration. For these reasons, we also leverage split-compilation [2]: the idea consists in pre-computing alternatives, and embedding in the program enough information to assist and drive a runtime system towards to the best solution.

#### 3.2.2. Software Adaptation

More than ever, software needs to adapt to its environment. In most cases, this environment remains unknown until runtime. This is already the case when one deploys an application to a cloud, or an "app" to mobile devices. The dilemma is the following: for maximum portability, developers should target the most general device; but for performance they would like to exploit the most recent and advanced hardware features. JIT compilers can handle the situation to some extent, but binary deployment requires dynamic binary rewriting. Our work has shown how SIMD instructions can be upgraded from SSE to AVX [3]. Many more opportunities will appear with diverse and heterogeneous processors, featuring various kinds of accelerators.

On shared hardware, the environment is also defined by other applications competing for the same computational resources. It becomes increasingly important to adapt to changing runtime conditions, such as the contention of the cache memories, available bandwidth, or hardware faults. Fortunately, optimizing at runtime is also an opportunity, because this is the first time the program is visible as a whole: executable and libraries (including library versions). Optimizers may also rely on dynamic information, such as actual input data, parameter values, etc. We have already developed a software platform [16] to analyze and optimize programs at runtime, and we started working on automatic dynamic parallelization of sequential code, and dynamic specialization.

We started addressing some of these challenges in ongoing projects such as Nano2017 PSAIC Collaborative research program with STMicroelectronics, as well as within the Inria Project Lab MULTICORE. The starting H2020 FET HPC project ANTAREX also addresses these challenges from the energy perspective. We further leverage our platform and initial results to address other adaptation opportunities. Efficient software adaptation requires expertise from all domains tackled by PACAP, and strong interaction between all team members is expected.

<sup>&</sup>lt;sup>0</sup>According to Dennard scaling, as transistors get smaller the power density remains constant, and the consumed power remains proportional to the area.

#### 3.2.3. Research directions in uniprocessor microarchitecture

Achieving high single-thread performance remains a major challenge even in the multicore era (Amdahl's law). The members of the PACAP project-team have been conducting research in uniprocessor microarchitecture research for about 20 years covering major topics including caches, instruction front-end, branch prediction, out-of-order core pipeline, and value prediction. In particular, in the recent years they have been recognized as world leaders in branch prediction [22][11] and in cache prefetching [9] and they have revived the forgotten concept of value prediction [14][13]. This research was supported by the ERC Advanced grant DAL (2011-2016) and also by Intel. We pursue research on achieving ultimate unicore performance. Below are several non-orthogonal directions that we have identified for mid-term research:

- 1. management of the memory hierarchy (particularly the hardware prefetching);
- 2. practical design of very wide issue execution core;
- 3. speculative execution.

#### Memory design issues:

Performance of many applications is highly impacted by the memory hierarchy behavior. The interactions between the different components in the memory hierarchy and the out-of-order execution engine have high impact on performance.

The last *Data Prefetching Contest* held with ISCA 2015 has illustrated that achieving high prefetching efficiency is still a challenge for wide-issue superscalar processors, particularly those featuring a very large instruction window. The large instruction window enables an implicit data prefetcher. The interaction between this implicit hardware prefetcher and the explicit hardware prefetcher is still relatively mysterious as illustrated by Pierre Michaud's BO prefetcher (winner of DPC2) [9]. The first objective of the research is to better understand how the implicit prefetching enabled by the large instruction window interacts with the L2 prefetcher.

The second objective of the research is related to the interaction of prefetching and virtual/physical memory. On real hardware, prefetching is stopped by page frontiers. The interaction between TLB prefetching (and on which level) and cache prefetching must be analyzed.

The prefetcher is not the only actor in the hierarchy that must be carefully controlled. Significant benefit can also be achieved through careful management of memory access bandwidth, particularly the management of spatial locality on memory accesses, both for reads and writes. The exploitation of this locality is traditionally handled in the memory controller. However, it could be better handled if larger temporal granularity was available. Finally, we also intend to continue to explore the promising avenue of compressed caches. In particular we recently proposed the skewed compressed cache [17]. It offers new possibility for efficient compression schemes.

#### Ultra wide-issue superscalar.

To effectively leverage memory level parallelism, one requires huge out-of-order execution structures as well as very wide issue superscalar processor. For the two past decades, implementing ever wider issue superscalar processor has been challenging. The objective of our research on the execution core is to explore (and revisit) directions to allow the design of a very wide-issue (8-to-16 way) out-of-order execution core while mastering its complexity (silicon area, hardware logic complexity, power/energy consumption).

The first direction that we are exploring is the use of clustered architecture [10]. Symmetric clustered organization allows to benefit from simpler bypass network, but induce large complexity on the issue queue. One remarkable finding of our study [10] is that, when considering two large clusters (e.g. 8-wide) steering large groups of consecutive instructions (e.g. 64  $\mu$ ops) to the same cluster is quite efficient. This opens opportunities to limit the complexity of the issue queues (monitoring fewer buses) and register files (fewer ports and physical registers) in the clusters, since not all results have to be forwarded to the other cluster.

The second direction that we are exploring is associated with the approach that we developed with Sembrant et al. [18]. It reduces the number of instructions waiting in the instruction queues for the applications benefiting from very large instruction windows. Instructions are dynamically classified as ready (independent from any long latency instruction) or non-ready, and as urgent (part of a dependency chain leading to a long latency instruction) or non-ready non-urgent instructions can be delayed until the long latency instruction has been executed; this allows to reduce the pressure on the issue queue. This proposition opens the opportunity to consider an asymmetric microarchitecture with a cluster dedicated to the execution of urgent instructions and a second cluster executing the non-urgent instructions. The microarchitecture of this second cluster could be optimized to reduce complexity and power consumption (smaller instruction queue, less aggressive scheduling...)

#### Speculative execution.

Out-of-order (OoO) execution relies on speculative execution that requires predictions of all sorts: branch, memory dependency, value...

The PACAP members have been major actors of the branch prediction research for the last 20 years; and their proposals have influenced the design of most of the hardware branch predictors in current microprocessors. We will continue to steadily explore new branch predictor designs as for instance [20].

In speculative execution, we have recently revisited value prediction (VP) which was a hot research topic between 1996 and 2002. However it was considered until recently that value prediction would lead to a huge increase in complexity and power consumption in every stage of the pipeline. Fortunately, we have recently shown that complexity usually introduced by value prediction in the OoO engine can be overcome [14][13][22][11]. First, very high accuracy can be enforced at reasonable cost in coverage and minimal complexity [14]. Thus, both prediction validation and recovery by squashing can be done outside the out-of-order engine, at commit time. Furthermore, we propose a new pipeline organization, EOLE ({Early | Out-of-order | Late} Execution), that leverages VP with validation at commit to execute many instructions outside the OoO core, in-order [13]. With EOLE, the issue-width in OoO core can be reduced without sacrificing performance, thus benefiting the performance of VP without a significant cost in silicon area and/or energy. In the near future, we will explore new avenues related to value prediction. These directions include register equality prediction and compatibility of value prediction with weak memory models in multiprocessors.

#### 3.2.4. Towards heterogeneous single-ISA CPU-GPU architectures

Heterogeneous single-ISA architectures have been proposed in the literature during the 2000's [55] and are now widely used in the industry (ARM big.LITTLE, NVIDIA 4+1...) as a way to improve power-efficiency in mobile processors. These architectures include multiple cores whose respective microarchitectures offer different trade-offs between performance and energy efficiency, or between latency and throughput, while offering the same interface to software. Dynamic task migration policies leverage the heterogeneity of the platform by using the most suitable core for each application, or even each phase of processing. However, these works only tune cores by changing their complexity. Energy-optimized cores are either identical cores implemented in a low-power process technology, or simplified in-order superscalar cores, which are far from state-of-the-art throughput-oriented architectures such as GPUs.

We investigate the convergence of CPU and GPU at both architecture and compilation levels.

#### Architecture.

The architecture convergence between Single Instruction Multiple Threads (SIMT) GPUs and multicore processors that we have been pursuing [7] opens the way for heterogeneous architectures including latency-optimized superscalar cores and throughput-optimized GPU-style cores, which all share the same instruction set. Using SIMT cores in place of superscalar cores will enable the highest energy efficiency on regular sections of applications. As with existing single-ISA heterogeneous architectures, task migration will not necessitate any software rewrite and will accelerate existing applications.

Compilers for emerging heterogeneous architectures.

Single-ISA CPU+GPU architectures will provide the necessary substrate to enable efficient heterogeneous processing. However, it will also introduce substantial challenges at the software and firmware level. Task placement and migration will require advanced policies that leverage both static information at compile time and dynamic information at run-time. We are tackling the heterogeneous task scheduling problem at the compiler level. As a first step, we are prototyping scheduling algorithms on existing multiple-ISA CPU+GPU architectures like NVIDIA Tegra X1.

#### 3.2.5. Real-time systems

Safety-critical systems (e.g. avionics, medical devices, automotive...) have so far used simple unicore hardware systems as a way to control their predictability, in order to meet timing constraints. Still, many critical embedded systems have increasing demand in computing power, and simple unicore processors are not sufficient anymore. General-purpose multicore processors are not suitable for safety-critical real-time systems, because they include complex micro-architectural elements (cache hierarchies, branch, stride and value predictors) meant to improve average-case performance, and for which worst-case performance is difficult to predict. The prerequisite for calculating tight WCET is a deterministic hardware system that avoids dynamic, time-unpredictable calculations at run-time.

Even for multi and manycore systems designed with time-predictability in mind (Kalray MPPA manycore architecture <sup>0</sup>, or the Recore manycore hardware <sup>0</sup>) calculating WCETs is still challenging. The following two challenges will be addressed in the mid-term:

- 1. definition of methods to estimate WCETs tightly on manycores, that smartly analyzes and/or controls shared resources such as buses, NoCs or caches;
- 2. methods to improve the programmability of real-time applications through automatic parallelization and optimizations from model-based designs.

#### 3.2.6. Fault Tolerance

Technology trends suggest that, in tomorrow's computing world, failures will become commonplace due to many factors, and the expected probability of failure will increase with scaling. While well-known approaches, such as error correcting codes, exist to recover from failures and provide fault-free chips, the exponential growth of the number of faults will make them unaffordable in the future. Consequently, other approaches such as fine-grained disabling and reconfiguration of hardware elements (e.g. individual functional units or cache blocks) will become economically necessary. We are going to enter a new era: functionally correct chips with variable performance among chips and throughout their lifetime [56].

Transient and permanent faults may be detected by similar techniques, but correcting them generally involves different approaches. We are primarily interested in permanent faults, even though we do not necessarily disregard transient faults (e.g. the TMR approach in the next paragraph addresses both kind of faults).

#### CPU.

Permanent faults can occur anywhere in the processor. The performance implications of faulty cells vary depending on how the array is used in a processor. Most of micro-architectural work aiming at assessing the performance implications of permanently faulty cells relies on simulations with random fault-maps. These studies are, therefore, limited by the fault-maps they use that may not be representative for the average and distributed performance. They also do not consider aging effect.

Considering the memory hierarchy, we have already studied [5] the impact of permanent faults on the average and worst-case performance based on analytical models. We will extend these models to cover other components and other designs, and to analyze the interaction between faulty components.

For identified critical hardware structures, such as the memory hierarchy, we will propose protection mechanisms by for instance using larger cells, or even by selecting a different array organization to mitigate the impact of faults.

<sup>&</sup>lt;sup>0</sup>http://www.kalrayinc.com

<sup>&</sup>lt;sup>0</sup>http://www.recoresystems.com/

Another approach to deal with faults is to introduce redundancy at the code level. We propose to consider static compilation techniques focusing on existing hardware. As an example, we plan to leverage SIMD extensions of current instruction sets to introduce redundancy in scalar code at minimum cost. With these instructions, it will be possible to protect the execution from both soft errors by using TMR (triple modular redundancy) with voters in the code itself, and permanent faults without the need of extra hardware support to deconfigure faulty functional units.

#### Reconfigurable Computing.

In collaboration with the CAIRN project-team, we propose to construct Coarse Grain Reconfigurable Architectures (CGRA) from a sea of basic arithmetic and memory elements organized into clusters and connected through a hierarchical interconnection network. These clusters of basic arithmetic operators (e.g. 8-bit arithmetic and logic units) would be able to be seamlessly configured to various accuracy and data types to adapt the consumed energy to application requirements taking advantage of approximate computations. We propose to add new kinds of error detection (and sometimes correction) directly at the operator level by taking advantage of the massive redundancy of the array. As an example, errors can be tracked and detected in a complex sequence of double floating-point operations by using a reduced-precision version of the same processing.

Such reconfigurable blocks will be driven by compilation techniques, in charge of computing checkpoints, detecting faults, and replaying computations when needed.

Dynamic compilation techniques will help better exploit faulty hardware, by allocating data and computations on correct resources. In case of permanent faults, we will provide a mechanism to reconfigure the hardware, for example by reducing the issue width of VLIW processors implemented in CGRA. Dynamic code generation (JIT compiler) will re-generate code for the new configuration, guaranteeing portability and optimal exploitation of the hardware.

#### 3.2.7. Power efficiency

PACAP addresses power-efficiency at several levels. First, we design static and split compilation techniques to contribute to the race for Exascale computing (the general goal is to reach  $10^{18}$  FLOP/s at less than 20 MW). Second, we focus on high-performance low-power embedded compute nodes. Within the ANR project Continuum, in collaboration with architecture and technology experts from LIRMM and the SME Cortus, we research new static and dynamic compilation techniques that fully exploit emerging memory and NoC technologies. Finally, in collaboration with the CAIRN project-team, we investigate the synergy of reconfigurable computing and dynamic code generation.

#### *Green and heterogeneous high-performance computing.*

Concerning HPC systems, our approach consists in mapping, runtime managing and autotuning applications for green and heterogeneous High-Performance Computing systems up to the Exascale level. One key innovation of the proposed approach consists of introducing a separation of concerns (where self-adaptivity and energy efficient strategies are specified aside to application functionalities) promoted by the definition of a Domain Specific Language (DSL) inspired by aspect-oriented programming concepts for heterogeneous systems. The new DSL will be introduced for expressing adaptivity/energy/performance strategies and to enforce at runtime application autotuning and resource and power management. The goal is to support the parallelism, scalability and adaptability of a dynamic workload by exploiting the full system capabilities (including energy management) for emerging large-scale and extreme-scale systems, while reducing the Total Cost of Ownership (TCO) for companies and public organizations.

#### High-performance low-power embedded compute nodes.

We will address the design of next generation energy-efficient high-performance embedded compute nodes. It focuses at the same time on software, architecture and emerging memory and communication technologies in order to synergistically exploit their corresponding features. The approach of the project is organized around three complementary topics: 1) compilation techniques; 2) multicore architectures; 3) emerging memory and communication technologies. PACAP will focus on the compilation aspects, taking as input the software-visible characteristics of the proposed emerging technology, and making the best possible use of the new features (non-volatility, density, endurance, low-power).

#### Hardware Accelerated JIT Compilation.

Reconfigurable hardware offers the opportunity to limit power consumption by dynamically adjusting the number of available resources to the requirements of the running software. In particular, VLIW processors can adjust the number of available issue lanes. Unfortunately, changing the processor width often requires recompiling the application, and VLIW processors are highly dependent of the quality of the compilation, mainly because of the instruction scheduling phase performed by the compiler. Another challenge lies in the high constraints of the embedded system: the energy and execution time overhead due to the JIT compilation must be carefully kept under control.

We started exploring ways to reduce the cost of JIT compilation targeting VLIW-based heterogeneous manycore systems. Our approach relies on a hardware/software JIT compiler framework. While basic optimizations and JIT management are performed in software, the compilation back-end is implemented by means of specialized hardware. This back-end involves both instruction scheduling and register allocation, which are known to be the most time-consuming stages of such a compiler.

#### 3.2.8. Security

Security is a mandatory concern of any modern computing system. Various threat models have led to a multitude of protection solutions. Members of PACAP already contributed, thanks to the HAVEGE [59] random number generator, and code obfuscating techniques (the obfuscating just-in-time compiler [54], or thread-based control flow mangling [58]).

We partner with security experts who can provide intuition, know-how and expertise, in particular in defining threat models, and assessing the quality of the solutions. Our background in compilation and architecture helps design more efficient and less expensive protection mechanisms.

We already have ongoing research directions related to security. SECODE (Secure Codes to Thwart Cyberphysical Attacks) is a project started January 2016, in collaboration with security experts from Télécom Paris Tech, Paris 8, Université Catholique de Louvain (Belgium), and University of Sabancı (Turkey). We also plan to partner with the Inria/CentraleSupelec CIDRE project-team to design a tainting technique based on a just-in-time compiler.

#### Compiler-based data protection.

We specify and design error correction codes suitable for an efficient protection of sensitive information in the context of Internet of Things (IoT) and connected objects. We partner with experts in security and codes to prototype a platform that demonstrates resilient software. PACAP's expertise is key to select and tune the protection mechanisms developed within the project, and to propose safe, yet cost-effective solutions from an implementation point of view.

#### JIT-based tainting.

Dynamic information flow control (DIFC, also known as *tainting*) is used to detect intrusions and to identify vulnerabilities. It consists in attaching metadata (called *taints* or *labels*) to information containers, and to propagate the taints when particular operations are applied to the containers: reads, writes, etc. The goal is then to guarantee that confidential information is never used to generate data sent to an untrusted container; conversely, data produced by untrusted entities cannot be used to update sensitive data.

The containers can be of various granularities: fine-grain approaches can deal with single variables, coarsergrain approaches consider a file as a whole. The CIDRE project-team has developed several DIFC monitors. kBlare is coarse-grain monitor in the Linux kernel. JBlare is a fine-grain monitor for Java applications. Finegrain monitors provide a better precision at the cost of a significant overhead in execution time.

Combining the expertise of CIDRE in DIFC with our expertise in JIT compilation will help design hybrid approaches. An initial static analysis of the program prior to installation or execution will feed information to a dynamic analyzer that propagates taints during just-in-time compilation.

# 4. Application Domains

#### 4.1. Any computer usage

The PACAP team is working on the fundamental technologies for computer science: processor architecture, performance-oriented compilation and guaranteed response time for real-time. The research results may have impacts on any application domain that requires high performance execution (telecommunication, multimedia, biology, health, engineering, environment...), but also on many embedded applications that exhibit other constraints such as power consumption, code size and guaranteed response time. Our research activity implies the development of software prototypes.

# 5. New Software and Platforms

#### **5.1. ATMI**

KEYWORDS: Analytic model - Chip design - Temperature

SCIENTIFIC DESCRIPTION: Research on temperature-aware computer architecture requires a chip temperature model. General purpose models based on classical numerical methods like finite differences or finite elements are not appropriate for such research, because they are generally too slow for modeling the timevarying thermal behavior of a processing chip.

We have developed an ad hoc temperature model, ATMI (Analytical model of Temperature in MIcroprocessors), for studying thermal behaviors over a time scale ranging from microseconds to several minutes. ATMI is based on an explicit solution to the heat equation and on the principle of superposition. ATMI can model any power density map that can be described as a superposition of rectangle sources, which is appropriate for modeling the microarchitectural units of a microprocessor.

FUNCTIONAL DESCRIPTION: ATMI is a library for modelling steady-state and time-varying temperature in microprocessors. ATMI uses a simplified representation of microprocessor packaging.

- Participant: Pierre Michaud
- Contact: Pierre Michaud
- URL: https://team.inria.fr/pacap/software/atmi/

#### **5.2. HEPTANE**

KEYWORDS: IPET - WCET - Performance - Real time - Static analysis - Worst Case Execution Time SCIENTIFIC DESCRIPTION: WCET estimation

Status: Registered with APP (Agence de Protection des Programmes). Available under GNU General Public License v3, with number IDDN.FR.001.510039.000.S.P.2003.000.10600.

The aim of Heptane is to produce upper bounds of the execution times of applications. It is targeted at applications with hard real-time requirements (automotive, railway, aerospace domains). Heptane computes WCETs using static analysis at the binary code level. It includes static analyses of microarchitectural elements such as caches and cache hierarchies.

For more information, please contact Damien Hardy or Isabelle Puaut.

FUNCTIONAL DESCRIPTION: In a hard real-time system, it is essential to comply with timing constraints, and Worst Case Execution Time (WCET) in particular. Timing analysis is performed at two levels: analysis of the WCET for each task in isolation taking account of the hardware architecture, and schedulability analysis of all the tasks in the system. Heptane is a static WCET analyser designed to address the first issue.

- Participants: Benjamin Lesage, Loïc Besnard, Damien Hardy, François Joulaud, Isabelle Puaut and Thomas Piquet
- Partner: Université de Rennes 1
- Contact: Isabelle Puaut
- URL: https://team.inria.fr/pacap/software/heptane/

#### 5.3. tiptop

KEYWORDS: Instructions - Cycles - Cache - CPU - Performance - HPC - Branch predictor SCIENTIFIC DESCRIPTION: Tiptop is written in C. It can take advantage of libncurses when available for pseudo-graphic display.

Performance, hardware counters, analysis tool.

Status: Registered with APP (Agence de Protection des Programmes). Available under GNU General Public License v2, with number IDDN.FR.001.450006.000.S.P.2011.000.10800. Current version is 2.3.1, released October 2017.

Tiptop has been integrated in major Linux distributions, such as Fedora, Debian, Ubuntu.

Tiptop is a new simple and flexible user-level tool that collects hardware counter data on Linux platforms (version 2.6.31+). The goal is to make the collection of performance and bottleneck data as simple as possible, including simple installation and usage. In particular, we stress the following points.

Installation is only a matter of compiling the source code. No patching of the Linux kernel is needed, and no special-purpose module needs to be loaded.

#### No privilege is required, any user can run tiptop

FUNCTIONAL DESCRIPTION: Today's microprocessors have become extremely complex. To better understand the multitude of internal events, manufacturers have integrated many monitoring counters. Tiptop can be used to collect and display the values from these performance counters very easily. Tiptop may be of interest to anyone who wants to optimise the performance of their HPC applications.

- Participant: Erven Rohou
- Contact: Erven Rohou
- URL: http://tiptop.gforge.inria.fr

#### **5.4. PADRONE**

KEYWORDS: Legacy code - Optimization - Performance analysis - Dynamic Optimization

FUNCTIONAL DESCRIPTION: Padrone is new platform for dynamic binary analysis and optimization. It provides an API to help clients design and develop analysis and optimization tools for binary executables. Padrone attaches to running applications, only needing the executable binary in memory. No source code or debug information is needed. No application restart is needed either. This is especially interesting for legacy or commercial applications, but also in the context of cloud deployment, where actual hardware is unknown, and other applications competing for hardware resources can vary. The profiling overhead is minimum.

- Participants: Emmanuel Riou and Erven Rohou
- Contact: Erven Rohou
- URL: https://team.inria.fr/alf/software/padrone

#### 5.5. If-memo

#### **KEYWORD:** Performance

SCIENTIFIC DESCRIPTION: We propose a linker based technique for enabling software memorizing of any dynamically linked pure function by function interception and we illustrate our framework using a set of computationally expensive pure functions - the transcendental functions. Our technique does not need the availability of source code and thus can even be applied to commercial applications as well as applications with legacy codes. As far as users are concerned, enabling memoization is as simple as setting an environment variable. Our framework does not make any specific assumptions about the underlying architecture or compiler too-chains, and can work with a variety of current architectures.

- Participants: Arjun Suresh and Erven Rohou
- Contact: Erven Rohou
- URL: https://team.inria.fr/alf/software/if-memo/

#### 5.6. Simty

KEYWORDS: RISC-V - Multi-threading - SIMT - FPGA - Softcore - GPU

FUNCTIONAL DESCRIPTION: Simty is a massively multi-threaded processor core that dynamically assembles SIMD instructions from scalar multi-thread code. It runs the RISC-V (RV32-I) instruction set. Unlike existing SIMD or SIMT processors like GPUs, Simty takes binaries compiled for general-purpose processors without any instruction set extension or compiler changes. Simty is described in synthesizable VHDL.

- Author: Sylvain Collange
- Contact: Sylvain Collange
- URL: https://gforge.inria.fr/projects/simty

#### 5.7. Barra

KEYWORDS: Performance - Computer architecture - Debug - Tesla ISA - GPU - Profiling - CUDA - HPC - Simulator - GPGPU

SCIENTIFIC DESCRIPTION: Research on throughout-oriented architectures demands accurate and representative models of GPU architectures in order to be able to evaluate new architectural ideas, explore design spaces and characterize applications. The Barra project is a simulator of the NVIDIA Tesla GPU architecture.

Barra builds upon knowledge acquired through micro-benchmarking, in order to provide a baseline model representative of industry practice. The simulator provides detailed statistics to identify optimization opportunities and is fully customizable to experiment ideas of architectural modifications. Barra incorporates both a functional model and a cycle-level performance model.

FUNCTIONAL DESCRIPTION: Barra is a Graphics Processing Unit (GPU) architecture simulator. It simulates NVIDIA CUDA programs at the assembly language level. Barra is a tool for research on computer architecture, and can also be used to debug, profile and optimize CUDA programs at the lowest level.

RELEASE FUNCTIONAL DESCRIPTION: Timing model Tesla-like architecture model Fermi-like architecture model New per-PC control-flow divergence management Simultaneous branch and warp interweaving Affine vector cache

- Participants: Alexandre Kouyoumdjian, David Defour, Fabrice Mouhartem and Sylvain Collange
- Partners: ENS Lyon UPVD
- Contact: Sylvain Collange
- URL: http://barra.gforge.inria.fr/

# 6. New Results

#### 6.1. Compiler, vectorization, interpretation

**Participants:** Erven Rohou, André Seznec, Sylvain Collange, Rabab Bouziane, Arif Ali Ana-Pparakkal, Stefano Cherubin, Byron Hawkins, Arif Ali Ana-Pparakkal, Imane Lasri, Kévin Le Bon.

#### 6.1.1. Improving sequential performance through memoization

Participants: Erven Rohou, Imane Lasri, André Seznec.

Many applications perform repetitive computations, even when properly programmed and optimized. Performance can be improved by caching results of pure functions, and retrieving them instead of recomputing a result (a technique called memoization).

We previously proposed [23] a simple technique for enabling software memoization of any dynamically linked pure function and we illustrate our framework using a set of computationally expensive pure functions – the transcendental functions.

A restriction of the proposed framework was that memoization was restricted only to dynamically linked functions and the functions must be determined beforehand. We extended this work, and we propose function memoization using a compile-time technique thus extending the scope of memoization to user defined functions as well as making it transparently applicable to any dynamically linked functions. Our compile-time technique allows static linking of memoization code and this increases the benefit due to memoization by leveraging the inlining capability for the memoization wrapper. Our compile-time analysis can also handle functions with pointer parameters , and we handle constants more efficiently. Instruction set support can also be considered, and we propose associated hardware leading to additional performance gain.

This work was presented at the Compiler Construction Conference 2017 [50]. It is also described in the PhD thesis of Arjun Suresh [24].

#### 6.1.2. Optimization in the Presence of NVRAM

Participants: Erven Rohou, Rabab Bouziane.

Beyond the fact of generating machine code, compilers play a critical role in delivering high performance, and more recently high energy efficiency. For decades, the memory technology of target systems has consisted in SRAM at cache level, and DRAM for main memory. Emerging non-volatile memories (NVMs) open up new opportunities, along with new design challenges. In particular, the asymmetric cost of read/write accesses calls for adjusting existing techniques in order to efficiently exploit NVMs. In addition, this technology makes it possible to design memories with cheaper accesses at the cost of lower data retention times. These features can be exploited at compile time to derive better data mappings according to the application and data retention characteristics. We reviewed a number of compile-time analysis and optimization techniques, and how they could apply to systems in presence of NVMs [37]. In particular, we consider the case of the reduction of the number of writes, and the analysis of variables lifetime for memory bank assignment of program variables.

Concerning the reduction of writes, we propose a fast evaluation of NVM integration at cache level, together with a compile-time approach for mitigating the penalty incurred by the high write latency of STT-RAM. We implement a code optimization in LLVM for reducing so-called *silent stores*, i.e., store instruction instances that write to memory values that were already present there. This makes our optimization portable over any architecture supporting LLVM. Then, we assess the possible benefit of such an optimization on the Rodinia benchmark suite through an analytic approach based on parameters extracted from the literature devoted to NVMs. This makes it possible to rapidly analyze the impact of NVMs on memory energy consumption. Reported results show up to 42 % energy gain when considering STT-RAM caches. This work is accepted for publication at RAPIDO'18 [38].

This research is done in collaboration with Abdoulaye Gamatié at LIRMM (Montpellier) within the context the the ANR project CONTINUUM.

#### 6.1.3. Dynamic Binary Optimization

Participants: Erven Rohou, Arif Ali Ana-Pparakkal, Kévin Le Bon, Byron Hawkins.

6.1.3.1. Dynamic Function Specialization

Participants: Erven Rohou, Arif Ali Ana-Pparakkal, Kévin Le Bon.

Compilers can do better optimization with the knowledge of run-time behavior of the program. *Function specialization* is a compilation technique that consists in optimizing the body of a function for specific values of an argument. Different versions of a function are created to deal with the most frequent values of the arguments, as well as the default case. Compilers can do a better optimization with the knowledge of run-time behaviour of the program. Static compilers, however, can hardly predict the exact value/behaviour of arguments, and even profiling collected during previous runs is never guaranteed to capture future behaviour. We propose a dynamic function specialization technique, that captures the actual values of arguments during execution of the program and, when profitable, creates specialized versions and include them at runtime. Our approach relies on dynamic binary rewriting. We present [36] the principles and implementation details of our technique, analyze sources of overhead, and present our results.

This research is done within the context of the Nano 2017 PSAIC collaborative project.

#### 6.1.3.2. Runtime Vectorization of Binary Programs

#### Participant: Erven Rohou.

In many cases, applications are not optimized for the hardware on which they run. Several reasons contribute to this unsatisfying situation, such as legacy code, commercial code distributed in binary form, or deployment on compute farms. In fact, backward compatibility of ISA guarantees only the functionality, not the best exploitation of the hardware. In this work, we focus on maximizing the CPU efficiency for the SIMD extensions.

We previously proposed [3] a binary-to-binary optimization framework where loops vectorized for an older version of the processor SIMD extension are automatically converted to a newer one. It is a lightweight mechanism that does not include a vectorizer, but instead leverages what a static vectorizer previously did. We showed that many loops compiled for x86 SSE can be dynamically converted to the more recent and more powerful AVX; as well as, how correctness is maintained with regards to challenges such as data dependencies and reductions. We obtained speedups in line with those of a native compiler targeting AVX.

We now focus on runtime vectorization of loops in binary codes that were not originally vectorized [29]. For this purpose, we use open source frameworks that we have tuned and integrated to

- 1. dynamically lift the x86 binary into the Intermediate Representation form of the LLVM compiler,
- 2. abstract hot loops in the polyhedral model,
- 3. use the power of this mathematical framework to vectorize them,
- 4. and finally compile them back into executable form using the LLVM Just-In-Time compiler.

In most cases, the obtained speedups are close to the number of elements that can be simultaneously processed by the SIMD unit. The re-vectorizer and auto-vectorizer are implemented inside a dynamic optimization platform; it is completely transparent to the user, does not require any rewriting of the binaries, and operates during program execution.

This work is done in collaboration with Philippe Clauss (Inria CAMUS), it is part of the PhD work of Nabil Hallou [26].

#### 6.1.4. Hardware/Software JIT Compiler

#### Participant: Erven Rohou.

Dynamic Binary Translation (DBT) is often used in hardware/software co-design to take advantage of an architecture model while using binaries from another one. The co-development of the DBT engine and of the execution architecture leads to architecture with special support to these mechanisms. We proposed [46] a hardware accelerated dynamic binary translation where the first steps of the DBT process are fully accelerated in hardware. Results showed that using our hardware accelerators leads to a speed-up of  $8\times$  and a cost in energy  $18\times$  lower, compared with an equivalent software approach.

Single ISA-Heterogeneous multi-cores such as the ARM big.LITTLE have proven to be an attractive solution to explore different energy/performance trade-offs. Such architectures combine Out of Order cores with smaller in-order ones to offer different power/energy profiles. They however do not really exploit the characteristics of workloads (compute-intensive vs. control dominated). In our recent work, we propose to enrich these architectures with runtime configurable VLIW cores, which are very efficient at compute-intensive kernels. To preserve the single ISA programming model, we resort to Dynamic Binary Translation, and use this technique to enable dynamic code specialization for Runtime Reconfigurable VLIWs cores. Our proposed DBT framework targets the RISC-V ISA, for which both OoO and in-order implementations exist. Our experimental results show that our approach can lead to best-case performance and energy efficiency when compared against static VLIW configurations.

This work has been accepted for publication at DATE 2018 [53].

This research is done in collaboration with Steven Derrien and Simon Rokicki from the CAIRN team.

#### 6.1.5. Customized Precision Computing

Participants: Erven Rohou, Stefano Cherubin, Imane Lasri.

Error-tolerating applications are increasingly common in the emerging field of real-time HPC. Proposals have been made at the hardware level to take advantage of inherent perceptual limitations, redundant data, or reduced precision input, as well as to reduce system costs or improve power efficiency. At the same time, works on floating-point to fixed-point conversion tools allow us to trade-off the algorithm exactness for a more efficient implementation. In this work [39], we aim at leveraging existing, HPC-oriented hardware architectures, while including in the precision tuning an adaptive selection of floating-and fixed-point arithmetic. Our proposed solution takes advantage of the application domain knowledge of the programmers by involving them in the first step of the interaction chain. We rely on annotations written by the programmer on the input file to know which variables of a computational kernel should be converted to fixed-point. The second stage replaces the floating-point variables in the kernel with fixed-point equivalents. It also adds to the original source code the utility functions to perform data type conversions from floating-point to fixed-point, and vice versa. The output of the second stage is a new version of the kernel source code which exploits fixed-point computation instead of floating-point computation. As opposed to typical custom-width hardware designs, we only rely on the standard 16-bit, 32-bit and 64-bit types. We also explore the impact of the fixedpoint representation on auto-vectorization. We discuss the effect of our solution in terms of time-to-solutions, error and energy-to-solution.

This is done within the context of the ANTAREX project in collaboration with Stefano Cherubin, and Giovanni Agosta from Politecnico di Milano, and Olivier Sentieys from the CAIRN team.

#### 6.1.6. SPMD Function Call Re-Vectorization

Participant: Sylvain Collange.

SPMD programming languages for SIMD hardware such as C for CUDA, OpenCL or ISPC have contributed to increase the programmability of SIMD accelerators and graphics processing units. However, SPMD languages still lack the flexibility offered by low-level SIMD programming on explicit vectors. To close this expressiveness gap while preserving the SPMD abstraction, we introduce the notion of Function Call Re-Vectorization (CREV). CREV allows changing the dimension of vectorization during the execution of an SPMD kernel, and exposes it as a nested parallel kernel call. CREV affords a programmability close to dynamic parallelism, a feature that allows the invocation of kernels from inside kernels, but at much lower cost. We defined a formal semantics of CREV, and implemented it on the ISPC compiler. To validate our idea, we have used CREV to implement some classic algorithms, including string matching, depth first search and Bellman-Ford, with minimum effort. These algorithms, once compiled by ISPC to Intel-based vector instructions, are as fast as state-of-the-art implementations, yet much simpler. As an example, our straightforward implementation of string matching beats the Knuth-Morris-Pratt algorithm by 12%. This work was presented at the ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP) 2017 [45].

This work was done in collaboration with Rubens Emilio and Fernando Pereira at UFMG, as part of the Inria PROSPIEL Associate Team.

#### 6.1.7. Qubit allocation for quantum circuit compilers

Participant: Sylvain Collange.

Quantum computing hardware is becoming a reality. For instance, IBM Research makes a quantum processor available in the cloud to the general public. The possibility of programming an actual quantum device has elicited much enthusiasm. Yet, quantum programming still lacks the compiler support that modern programming languages enjoy today. To use universal quantum computers like IBM's, programmers must design low-level circuits. In particular, they must map logical qubits into physical qubits that need to obey connectivity constraints. This task resembles the early days of programming, in which software was built in machine languages. We have formally introduced the qubit allocation problem and provided an exact solution to it. This optimal algorithm deals with the simple quantum machinery available today; however, it cannot scale up to the more complex architectures scheduled to appear. Thus, we also provide a heuristic solution to qubit allocation, which is faster than the current solutions already implemented to deal with this problem.

This paper is accepted for publication at the Code Generation and Optimization (CGO) conference [49].

This work was done in collaboration with Vinícius Fernandes dos Santos, Fernando Pereira and Marcos Yukio Siraichi at UFMG, Brazil.

#### **6.2.** Processor Architecture

**Participants:** Pierre Michaud, Sylvain Collange, Erven Rohou, André Seznec, Biswabandan Panda, Fernando Endo, Kleovoulos Kalaitzidis, Daniel Rodrigues Carvalho, Anita Tino.

Processor, cache, locality, memory hierarchy, branch prediction, multicore, power, temperature

#### 6.2.1. Microarchitecture

6.2.1.1. Bayesian TAGE predictors

Participant: Pierre Michaud.

The TAGE conditional branch predictor, introduced by André Seznec and Pierre Michaud in 2006, is the most storage-efficient branch predictor known today [19]. André Seznec has won the last four branch prediction championships, each time with a TAGE-based predictor. However, since 2006, the improvements in prediction accuracy have been relatively modest and were mostly obtained at the cost of increased hardware complexity. In particular, André Seznec added a Statistical Corrector to TAGE to correct some of its deficiencies [21]. This may be an indication that our understanding of TAGE is not complete and that further accuracy gains are waiting to be discovered. The problem tackled by the statistical corrector is that of cold counters: a TAGElike predictor constantly allocate new entries, erasing the branch history information stored in the up-down counters of the overwritten entries. TAGE mitigates this problem by using the confidence level of the updown counter and a meta-predictor. However, fundamentally, the information on the degree of coldness of the up-down counter is not available in TAGE. Therefore we propose to replace the up-down counter with a dual-counter counting separately taken and not-taken occurrences. Replacing the up-down counter with a dual-counter requires to redefine prediction confidence estimation. We found that a Bayesian formula, namely Laplace's rule of succession, provides effective confidence estimation. We also discovered a method, based on the dual-counter, for reducing the number of allocations. By combining these new findings, we devised a new TAGE-like predictor called BATAGE, more accurate than TAGE, making external statistical correction superfluous. As of December 2017, this work is in the process of being submitted to a journal.

#### 6.2.1.2. Interactions Between Value Prediction and Compiler Optimizations Participants: André Seznec, Fernando Endo.

Increasing instruction-level parallelism is regaining attractiveness within the microprocessor industry. The EOLE microarchitecture [13] and D-VTAGE value predictor [14] were recently introduced to solve practical issues of value prediction (VP). In particular, they remove the most significant difficulties that forbade an effective VP hardware. In [28], we present a detailed evaluation of the potential of VP in the context of EOLE/D-VTAGE and different compiler options. Our study shows that if no single general rule always applies – more optimization might sometimes leads to more performance – unoptimized codes often gets a large benefit from the prediction of redundant loads.

6.2.1.3. Prefetch Management on Multicore Systems Participants: André Seznec, Biswabandan Panda.

In multi-core systems, an application's prefetcher can interfere with the memory requests of other applications using the shared resources, such as last level cache and memory bandwidth. Towards this end, we propose a solution to manage prefetching in multi-core systems [32]. In particular, we make two fundamental observations: First, a strong positive correlation exists between the accuracy of a prefetcher and the amount of prefetch requests it generates relative to an application's total (demand and prefetch) requests. Second, a strong positive correlation exists between the ratio of total prefetch to demand requests and the ratio of average last level cache miss service times of demand to prefetch requests. In [32], we propose Band-pass prefetching a simple and low-overhead mechanism to effectively manage prefetchers in multi-core systems that builds on those two observations. Our solution consists of local and global prefetcher aggressiveness control components, which altogether, control the flow of prefetch requests between a range of prefetch to demand requests ratios.

#### 6.2.1.4. Managing Shared Last Level Caches in Large Multicores Participant: André Seznec.

Multi-core processors employ shared Last Level Caches (LLC). This trend continuewith large multi-core processors (16 cores and beyond) as well. At the same time, the associativity of LLC tends to remain in the order of sixteen. Consequently, with large multicore processors, the number of applications or threads that share the LLC becomes larger than the associativity of the cache itself. LLC management policies have been extensively studied for small scale multi-cores (4 to 8 cores) and associativity degree in the 16 range. However, the impact of LLC management on large multi-cores is essentially unknown, in particular when the associativity degree is smaller than the number of applications. In [33], we introduce Adaptive Discrete and deprioritized Application PrioriTization (ADAPT), an LLC management policy addressing the large multi-cores where the LLC associativity degree is smaller than the number of applications. ADAPT builds on the use of the Footprint-number metric. We propose a monitoring mechanism that dynamically samples cache sets to estimate the Footprint-number of applications and classifies them into discrete (distinct and more than two) priority buckets. The cache replacement policy leverages this classification and assigns priorities to cache lines of applications during cache replacement operations. We further find that deprioritizing certain applications during cache replacement is beneficial to the overall performance.

6.2.1.5. Augmenting superscalar architecture for efficient many-thread parallel execution **Participants:** Sylvain Collange, André Seznec.

Threads of Single-Program Multiple-Data (SPMD) applications often exhibit very similar control flows, i.e. they execute the same instructions on different data. We propose the Dynamic Inter-Thread Vectorization Architecture (DITVA) to leverage this implicit data-level parallelism in SPMD applications by assembling dynamic vector instructions at runtime. DITVA extends an in-order SMT processor with SIMD units with an inter-thread vectorization execution mode. In this mode, multiple scalar threads running in lockstep share a single instruction stream and their respective instruction instances are aggregated into SIMD instructions. To balance thread-and data-level parallelism, threads are statically grouped into fixed-size independently scheduled warps. DITVA leverages existing SIMD units and maintains binary compatibility with existing CPU
architectures. Our evaluation on the SPMD applications from the PARSEC and Rodinia OpenMP benchmarks shows that a 4-warp  $\times$  4-lane 4-issue DITVA architecture with a realistic bank-interleaved cache achieves  $1.55 \times$  higher performance than a 4-thread 4-issue SMT architecture with AVX instructions while fetching and issuing 51 % fewer instructions, achieving an overall 24 % energy reduction.

This work has been accepted for publication in the Journal of Parallel and Distributed Computing [30]. It was done in collaboaration with Sajith Kalathingal and Bharath Swamy from Intel Bangalore (India).

6.2.1.6. Generalizing the SIMT execution model to general-purpose instruction sets **Participant:** Sylvain Collange.

The *Single Instruction, Multiple Threads* (SIMT) execution model as implemented in NVIDIA Graphics Processing Units (GPUs) associates a multi-thread programming model with an SIMD execution model [57]. It combines the simplicity of scalar code from the programmer's and compiler's perspective with the efficiency of SIMD execution units at the hardware level. However, current SIMT architectures demand specific instruction sets. In particular, they need specific branch instructions to manage thread divergence and convergence. Thus, SIMT GPUs have remained incompatible with traditional general-purpose CPU instruction sets.

We designed Simty, an SIMT processor proof of concept that lifts the instruction set incompatibility between CPUs and GPUs. Simty is a massively multi-threaded processor core that dynamically assembles SIMD instructions from scalar multi-thread code. It runs the RISC-V (RV32-I) instruction set. Unlike existing SIMD or SIMT processors like GPUs, Simty takes binaries compiled for general-purpose processors without any instruction set extension or compiler changes. Simty is described in synthesizable RTL. A FPGA prototype validates its scaling up to 2048 threads per core with 32-wide SIMD units.

The Simty architecture was presented at the First Workshop on Computer Architecture Research with RISC-V (CARRV 2017) [40].

Both conventional and generalized SIMT architectures like Simty use hardware or software mechanisms to keep track of control-flow divergence and convergence among threads. A new class of such algorithms is gaining popularity in the literature in the last few years. We presented a new classification of these techniques based on their common characteristic, namely traversals of the control-flow graph based on lists of paths. We compared the implementation cost on an FPGA of path lists and per-thread program counters within the Simty processor. The sorted list enables significantly better scaling starting from 8 threads per warp.

This work was presented in French in Conférence d'informatique en Parallélisme, Architecture et Système (ComPAS) [51] and is available in English as a technical report [52].

6.2.1.7. Toward out-of-order SIMT microarchitecture

Participants: Sylvain Collange, Anita Tino.

Prior work highlights the continued importance of maintaining adequate sequential performance within throughput-oriented cores [60]. Out-of-order superscalar architectures as used in high-performance CPU cores can meet such demand for single-thread performance. However, GPU architectures based on SIMT have been limited so far to in-order execution because of a major scientific obstacle: the partial dependencies between instructions that SIMT execution induces thwart register renaming. This ongoing project is seeking to generalize out-of-order execution to SIMT architectures. In particular, we revisit register renaming techniques originally proposed for predicate conversion to support partial register updates efficiently. Out-of-order dynamic vectorization holds the promise to close the CPU-GPU design space by enabling low-latency, high-throughput design points.

#### 6.3. WCET estimation and optimization

**Participants:** Isabelle Puaut, Damien Hardy, Viet Anh Nguyen, Benjamin Rouxel, Sébastien Martinez, Erven Rohou, Imen Fassi, Loïc Besnard, Stefanos Skalistis.

#### 6.3.1. WCET estimation for many core processors

Participants: Viet Anh Nguyen, Damien Hardy, Sébastien Martinez, Isabelle Puaut, Benjamin Rouxel.

#### 6.3.1.1. Optimization of WCETs by considering the effects of local caches

The overall goal of this research is to define WCET estimation methods for parallel applications running on many-core architectures, such as the Kalray MPPA machine.

Some approaches to reach this goal have been proposed, but they assume the mapping of parallel applications on cores already done. Unfortunately, on architectures with caches, task mapping requires a priori known WCETs for tasks, which in turn requires knowing task mapping (i.e., co-located tasks, co-running tasks) to have tight WCET bounds. Therefore, scheduling parallel applications and estimating their WCET introduce a chicken and egg situation.

We have addressed this issue by developing both optimal and heuristic techniques for solving the scheduling problem, whose objective is to minimize the WCET of a parallel application. Our proposed static partitioned non-preemptive mapping strategies address the effect of local caches to tighten the estimated WCET of the parallel application. Experimental results obtained on real and synthetic parallel applications show that colocating tasks that reuse code and data improves the WCET by 11 % on average for the optimal method and by 9 % on average for the heuristic method [35].

This research is part of the PIA Capacités project.

#### 6.3.1.2. Accounting for shared resource contentions to minimize WCETs

Accurate WCET analysis for multi-cores is known to be challenging, because of concurrent accesses to shared resources, such as communication through busses or Networks on Chips (NoC). Since it is impossible in general to guarantee the absence of resource conflicts during execution, current WCET techniques either produce pessimistic WCET estimates or constrain the execution to enforce the absence of conflicts, at the price of a significant hardware under-utilization. In addition, the large majority of existing works consider that the platform workload consists of independent tasks. As parallel programming is the most promising solution to improve performance, we envision that within only a few years from now, real-time workloads will evolve toward parallel programs. The WCET behavior of such programs is challenging to analyze because they consist of *dependent* tasks interacting through complex synchronization/communication mechanisms.

In a first work (thesis of Benjamin Rouxel), we proposed techniques that account for interferences to access shared ressources, in order to minimize the WCET of parallel applications. An optimal and a heuristic method are proposed to map and schedule tasks on multi-cores. These methods take the structure of applications (synchronizations/communications) into consideration to tightly identify shared resource interferences and consequently tighten WCET estimates. Our heuristic improves by 19% the overall WCET compared to a worst-case contention baseline [47], [31].

In a second study [44], we have studied the gain that could be obtained on an initially produced time-triggered non-preemptive schedule, by the introduction of slack time, in order to avoid interference between tasks. The introduction of slack time is performed using an optimal technique using Integer Linear Programming (ILP), to evaluate how much at best can be gained. Experimental results using synthetic task graphs and a Kalray-like architecture with round-robin bus arbitration show that avoiding contention reduces WCETs, albeit by a small percentage. The highest reductions are observed on applications with the highest memory demand, and when the application is scheduled on the highest number of cores.

This work is performed in cooperation with Steven Derrien from the CAIRN research group and is part of the ARGO H2020 project.

#### 6.3.1.3. WCET-Aware Parallelization of Model-Based Applications for Multi-Cores

Parallel architectures are nowadays no longer confined to the domain of high performance computing, they are also increasingly used in embedded time-critical systems.

The ongoing ARGO H2020 project provides a programming paradigm and associated tool flow to exploit the full potential of architectures in terms of development productivity, time-to-market, exploitation of the platform computing power and guaranteed real-time performance. In [41] we give an overview of the objectives of ARGO and explore the challenges introduced by our approach.

#### 6.3.2. WCET estimation tool and benchmarks

Participants: Damien Hardy, Isabelle Puaut, Benjamin Rouxel, Loïc Besnard.

Estimation of worst-case execution times (WCETs) is required to validate the temporal behavior of hard real time systems. Heptane is an open-source software program that estimates upper bounds of execution times on MIPS and ARM v7 architectures, offered to the WCET estimation community to experiment new WCET estimation techniques. The software architecture of Heptane was designed to be as modular and extensible as possible to facilitate the integration of new approaches. In [42], we present the current status of Heptane, give information on the analyses it implements, as well as how to use it and extend it.

We all had quite a time to find non-proprietary architecture-independent exploitable parallel benchmarks for Worst-Case Execution Time (WCET) estimation and real-time scheduling. However, there is no consensus on a parallel benchmark suite, when compared to the single-core era and the Mälardalen benchmark suite. In [48] we bridge part of this gap, by presenting a collection of benchmarks with the following good properties: (i) easily analyzable by static WCET estimation tools (written in structured C language, in particular neither goto nor dynamic memory allocation, containing flow information such as loop bounds); (ii) independent from any particular run-time system (MPI, OpenMP) or real-time operating system. Each benchmark is composed of the C source code of its tasks, and an XML description describing the structure of the application (tasks and amount of data exchanged between them when applicable). Each benchmark can be integrated in a full end-to-end empirical method validation protocol on multi-core architecture. This proposed collection of benchmarks is derived from the well known StreamIT benchmark suite and will be integrated in the TACleBench suite in a near future.

#### 6.4. Security

Participants: Erven Rohou, Damien Hardy, Nicolas Kiss.

Physical attacks represent a very important threat in the context of embedded systems: these attacks try to recover cryptographic keys by exploiting the physical behavior of the device. They can either be passive (e.g. by monitoring the power consumption of the device) or active (e.g. by injecting errors to reveal or deduce sensitive data).

One family of countermeasures to protect against those passive attacks (also known as *side-channel* attacks) is called masking. The principle is to "hide" data with masks so that internal values used in computations can not be predicted with the behavior observed. We modified the LLVM compiler (version 3.8) to automatically insert masking coutermeasures into the code at compile-time. Our modification works at intermediate level (IR level), this way we can perform low-level transformations (e.g. memory allocation, instructions replacement) while covering most of the architectures used in the embedded world.

The main innovation of this work is the generic approach used for the transformation and thus, the ability to easily change the masking scheme without modifying the compiler internal code. We introduced a way to describe in high-level language (C/C++) the masking operations independently in what we call "primitives". With this technique, we implemented "Boolean Masking" and we tested the efficiency on an embedded implementation of AES. After measuring the electromagnetic emissions of 20,000 executions, we performed a Correlation Power Analysis (CPA) and results have shown that the countermeasure is correctly applied. Hence, it is not possible anymore to recover the cryptographic key with this type of attack.

This work is done in the context of the SECODE CHIST-ERA project.

### 7. Bilateral Contracts and Grants with Industry

### 7.1. Bilateral Contracts with Industry

7.1.1. Nano 2017 PSAIC

Participants: Arif Ali Ana-Pparakkal, Erven Rohou.

Nano 2017 PSAIC is a collaborative R&D program involving Inria and STMicroelectronics. The PSAIC (Performance and Size Auto-tuning through Iterative Compilation) project concerns the automation of program optimization through the combination of several tools and techniques such as: compiler optimization, profiling, trace analysis, iterative optimization and binary analysis/rewriting. For any given application, the objective is to devise through a fully automated process a compiler profile optimized for performance and code size. For this purpose, we are developing instrumentation techniques that can be focused and specialized to a specific part of the application aimed to be monitored.

The project involves the Inria teams PACAP, AriC, CAMUS and CORSE. PACAP contributes program analyses at the binary level, as well as binary transformations. We will also study the synergy between static (compiler-level) and dynamic (run-time) analyses.

#### 7.2. Bilateral Grants with Industry

#### 7.2.1. Intel research grant INTEL2014-8957

Participants: André Seznec, Biswabandan Panda, Fernando Endo.

Intel is supporting the research of the PACAP project-team on "Mixing branch and value prediction to enable high sequential performance".

#### 7.2.2. Intel research grant INTEL2016-11174

Participants: André Seznec, Pierre Michaud, Kleovoulos Kalaitzidis, Niloofar Charmchi.

Intel is supporting the research of the PACAP project-team on "Design tradeoffs for extreme cores".

### 8. Partnerships and Cooperations

#### 8.1. Regional Initiatives

#### 8.1.1. Britanny region fellowship

Participants: Niloofar Charmchi, André Seznec.

The Britanny Region is partially funding a Ph.D. fellowship for Niloofar Charmchi on the topic "Hardware prefetching and related issues".

#### 8.2. National Initiatives

#### 8.2.1. Capacités: Projet "Investissement d'Avenir", 1/11/14 to 31/01/2018

Participants: Damien Hardy, Isabelle Puaut, Viet Anh Nguyen, Sébastien Martinez.

The project objective is to develop a hardware and software platform based on manycore architectures, and to demonstrate the relevance of these manycore architectures (and more specifically the Kalray manycore) for several industrial applications. The Kalray MPPA manycore architecture is currently the only one able to meet the needs of embedded systems simultaneously requiring high performance, lower power consumption, and the ability to meet the requirements of critical systems (low latency I/O, deterministic processing times, and dependability).

The project partners are Kalray (lead), Airbus, Open-Wide, Safran Sagem, IS2T, Real Time at Work, Dassault Aviation, Eurocopter, MBDA, ProbaYes, IRIT, Onera, Verimag, Inria, Irisa, Tima and Armines.

#### 8.2.2. Zero Power Computing Systems (ZEP): Inria Project Lab, 2017–2020

Participant: Erven Rohou.

This proposal addresses the issue of designing tiny wireless, batteryless, computing objects, harvesting energy in the environment. The energy level harvested being very low, very frequent energy shortages are expected. In order for the new system to maintain a consistent state, it will be based on a new architecture embedding non-volatile RAM (NVRAM). In order to benefit from the hardware innovations related to energy harvesting and NVRAM, software mechanisms will be designed. On the one hand, a compilation pass will compute a worst-case energy consumption. On the other hand, dedicated runtime mechanisms will allow:

- 1. to manage efficiently and correctly the NVRAM-based hardware architecture;
- 2. to use energy intelligently, by using the worst-case energy consumption.

The ZEP project gathers four Inria teams that have a scientific background in architecture, compilation, operating systems together with the CEA Lialp and Lisan laboratories of CEA LETI & LIST. The main application target is Internet of Things (IoT).

#### 8.2.3. ANR Continuum 2015–2019

Participants: Erven Rohou, Rabab Bouziane.

The CONTINUUM project aims to address the energy-efficiency challenge in future computing systems by investigating a design continuum for compute nodes, which seamlessly goes from software to technology levels via hardware architecture. Power saving opportunities exist at each of these levels, but the real measurable gains will come from the synergistic focus on all these levels as considered in this project. Then, a cross-disciplinary collaboration is promoted between computer science and microelectronics, to achieve two main breakthroughs: i) combination of state-of-the-art heterogeneous adaptive embedded multicore architectures with emerging communication and memory technologies and, ii) power-aware dynamic compilation techniques that suitably match such a platform.

Continuum started on Oct 1st 2015. Partners are LIRMM and Cortus SAS.

#### 8.2.4. ANR W-SEPT 2012-2017

Participants: Isabelle Puaut, Erven Rohou.

Critical embedded systems are generally composed of repetitive tasks that must meet drastic timing constraints, such as termination deadlines. Providing an upper bound of the worst-case execution time (WCET) of such tasks at design time is thus necessary to prove the correctness of the system. Static WCET estimation methods, although safe, may produce largely over-estimated values. The objective of the project is to produce tighter WCET estimates by discovering and transforming flow information at all levels of the software design process, from high level-design models (e.g. Scade, Simulink) down to binary code.

The ANR W-SEPT project partners are Verimag Grenoble, IRIT Toulouse, Inria Rennes. A case study is provided by Continental Toulouse.

#### 8.3. European Initiatives

#### 8.3.1. FP7 & H2020 Projects

#### 8.3.1.1. ANTAREX

Participants: Erven Rohou, Imane Lasri.

Title: Auto-Tuning and Adaptivity appRoach for Energy efficient exascale HPC Systems Program: H2020 Duration: September 2015 - September 2018 Coordinator: Politecnico di Milano, Italy (POLIMI) Partners: Consorzio Interuniversitario Cineca (Italy) Dompé Farmaceutici Spa (Italy) Eidgenoessische Technische Hochschule Zürich (Switzerland) Vysoka Skola Banska - Technicka Univerzita Ostrava (Czech Republic) Politecnico di Milano (Italy) Sygic As (Slovakia) Universidade do Porto (Portugal)

Inria contact: Erven Rohou

Energy-efficient heterogeneous supercomputing architectures need to be coupled with a radically new software stack capable of exploiting the benefits offered by the heterogeneity at all the different levels (supercomputer, job, node) to meet the scalability and energy efficiency required by Exascale supercomputers. ANTAREX will solve these challenging problems by proposing a disruptive holistic approach spanning all the decision layers composing the supercomputer software stack and exploiting effectively the full system capabilities (including heterogeneity and energy management). The main goal of the ANTAREX project is to provide a breakthrough approach to express application self-adaptivity at design-time and to runtime manage and autotune applications for green and heterogenous High Performance Computing (HPC) systems up to the Exascale level.

#### 8.3.1.2. Eurolab-4-HPC

#### Participant: André Seznec.

Title: EuroLab-4-HPC: Foundations of a European Research Center of Excellence in High Performance Computing Systems

Program: H2020

Duration: September 2015 - September 2017

Coordinator: Chalmers Tekniska Hoegskola AB

Partners:

Barcelona Supercomputing Center - Centro Nacional de Supercomputacion (Spain) Chalmers Tekniska Hoegskola (Sweden)

École Polytechnique Federale de Lausanne (Switzerland)

Foundation for Research and Technology Hellas (Greece)

Universität Stuttgart (Germany)

Rheinisch-Westfaelische Technische Hochschule Aachen (Germany)

Technion - Israel Institute of Technology (Israel)

Universität Augsburg (Germany)

The University of Edinburgh (United Kingdom)

Universiteit Gent (Belgium)

The University of Manchester (United Kingdom)

Inria contact: Albert Cohen (Inria Paris)

Europe has built momentum in becoming a leader in large parts of the HPC ecosystem. It has brought together technical and business stakeholders from application developers via system software to exascale systems. Despite such gains, excellence in high performance computing systems is often fragmented and opportunities for synergy missed. To compete internationally, Europe must bring together the best research groups to tackle the longterm challenges for HPC. These typically cut across layers, e.g., performance, energy efficiency and dependability, so excellence in research must target all the layers in the system stack. The EuroLab-4-HPC project's bold overall goal is to build connected and sustainable leadership in high-performance computing systems by bringing together the different and leading performance oriented communities in Europe, working across all layers of the system stack and, at the same time, fueling new industries in HPC.

#### 8.3.1.3. ARGO

Participants: Isabelle Puaut, Damien Hardy, Imen Fassi.

Title: Argo: WCET-Aware Parallelization of Model-Based Applications for Heterogeneous Parallel Systems Program: H2020 Type: RIA Duration: Jan 2016 - Dec 2018 Coordinator: Karlsruher Institut für Technologie (KIT) Université Rennes I contact: Steven Derrien Partners: Karlsruher Institut für Technologie (KIT) SCILAB enterprises SAS Recore Systems BV Université de Rennes I Technologiko Ekpaideftiko Idryma (TEI) Dytikis Elladas Absint GmbH Deutsches Zentrum für Luft - und Raumfahrt EV Fraunhofer

Increasing performance and reducing costs, while maintaining safety levels and programmability are the key demands for embedded and cyber-physical systems in European domains, e.g. aerospace, automation, and automotive. For many applications, the necessary performance with low energy consumption can only be provided by customized computing platforms based on heterogeneous manycore architectures. However, their parallel programming with time-critical embedded applications suffers from a complex toolchain and programming process. Argo (WCET-Aware PaRallelization of Model-Based Applications for HeteroGeneOus Parallel Systems) will address this challenge with a holistic approach for programming heterogeneous multi- and many-core architectures using automatic parallelization of model-based real-time applications. Argo will enhance WCET-aware automatic parallelization by a crosslayer programming approach combining automatic tool-based and user-guided parallelization to reduce the need for expertise in programming parallel heterogeneous architectures. The Argo approach will be assessed and demonstrated by prototyping comprehensive time-critical applications from both aerospace and industrial automation domains on customized heterogeneous many-core platforms.

Argo also involves Steven Derrien and Angeliki Kritikakou from the CAIRN team.

#### 8.3.1.4. HiPEAC4 NoE

Participants: Pierre Michaud, Erven Rohou, André Seznec.

P. Michaud, A. Seznec and E. Rohou are members of the European Network of Excellence HiPEAC4.

HiPEAC4 addresses the design and implementation of high-performance commodity computing devices in the 10+ year horizon, covering both the processor design, the optimizing compiler infrastructure, and the evaluation of upcoming applications made possible by the increased computing power of future devices.

### 8.4. International Initiatives

#### 8.4.1. ANR CHIST-ERA SECODE 2016-2018

Participants: Nicolas Kiss, Damien Hardy, Erven Rohou.

In this project, we specify and design error correction codes suitable for an efficient protection of sensitive information in the context of Internet of Things (IoT) and connected objects. Such codes mitigate passive attacks, like memory disclosure, and active attacks, like stack smashing. The innovation of this project is to leverage these codes for protecting against both cyber and physical attacks. The main advantage is a full coverage of attacks of the connected embedded systems, which is considered as a smart connected device and also a physical device. The outcome of the project is first a method to generate and execute cyber-resilient software, and second to protect data and its manipulation from physical threats like side-channel attacks. Theses results are demonstrated by using a smart sensor application with hardened embedded firmware and tamper-proof hardware platform.

Partners are Télécom Paris Tech, Université Paris 8, Sabancı Üniversitesi (Turkey), and Université Catholique de Louvain (Belgium).

#### 8.4.2. PHC IMHOTEP

Participant: Erven Rohou.

Title: Thoth – An Automatic Dynamic Binary Parallelisation System

International Partner (Institution - Laboratory - Researcher):

Egypt-Japan University of Science and Technology - Prof. Ahmed ElMahdy.

Dates: 2016-2017

With the current global trend towards utilizing cloud computing and smart devices, executing the same application across becomes a necessity. Moreover, parallelism is now abundant with various forms that include thread- and data-parallel execution models. Such diversity in ISA and explicit parallelism makes software development cost prohibitive, especially for natively optimized binaries. This project leverages dynamic binary translation technology to provide for exploiting the underlying parallel resources without the need of having the source code of the application. In particular the project integrates low overhead dynamic profiling, novel OSR parallel de-optimization and a retargetable parallelization modules to allow for dynamic parallelization of binaries.

#### 8.4.3. Inria Associate Teams Not Involved in an Inria International Labs

#### 8.4.3.1. PROSPIEL

Participant: Sylvain Collange.

Title: Profiling and specialization for locality

International Partner (Institution - Laboratory - Researcher):

Universidade Federal de Minas Gerais (Brazil) - DCC - Fernando Magno Quintão Pereira

Start year: 2015

See also: https://team.inria.fr/pacap/prospiel/

The PROSPIEL project aims at optimizing parallel applications for high performance on new throughput-oriented architectures: GPUs and many-core processors. Traditionally, code optimization is driven by a program analysis performed either statically at compile-time, or dynamically at run-time. Static program analysis is fully reliable but often over-conservative. Dynamic analysis provides more accurate data, but faces strong execution time constraints and does not provide any guarantee. By combining profiling-guided specialization of parallel programs with runtime checks for correctness, PROSPIEL seeks to capture the advantages of both static analysis and dynamic analysis. The project relies on the polytope model, a mathematical representation for parallel loops, as a theoretical foundation. It focuses on analyzing and optimizing performance aspects that become increasingly critical on modern parallel computer architectures: locality and regularity.

#### **8.5. International Research Visitors**

Prof. Ahmed ElMahdy, from the Egypt-Japan University of Science and Technology (E-JUST), Alexandria, Egypt, visited PACAP for two weeks in September, in the context of the project PHC IMHOTEP.

#### 8.5.1. Visits of International Scientists

8.5.1.1. Internships

Stefano Cherubin, PhD student at Politecnico di Milano for one month in Mar 2017, within the context of the ANTAREX H2020 project.

Andrei Rimsa Alvares, PhD at UFMG and Assistant Professor at CEFET-MG, 1 month from January 6 to February 5, 2017, PROSPIEL Associate Team.

### 9. Dissemination

#### 9.1. Promoting Scientific Activities

#### 9.1.1. Scientific Events Organisation

9.1.1.1. Member of the Organizing Committees

A. Seznec is member of the ACM/IEEE ISCA symposium steering committee.

#### 9.1.2. Scientific Events Selection

9.1.2.1. Chair of Conference Program Committees

Isabelle Puaut is Program Chair of the 2017 IEEE Real-Time Systems Symposium (RTSS), held in Paris, France. She is general Chair of RTSS 2018, to be held in Nashville, Tennessee.

9.1.2.2. Member of the Conference Program Committees

Sylvain Collange was PC member of Compas'2017.

Pierre Michaud was a member of the program committees of the ISCA 2017 and MICRO 2017 conferences.

Isabelle Puaut is member of the program committees of the Euromicro Conference on Real Time Systems (ECRTS) 2017 and 2018, the IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS) 2017 and 2018.

André Seznec is a member of IEEE Micro 2018 Top Picks selection committee.

André Seznec was a member of the IEEE Micro 2017 and SAMOS 2017 conference program committee.

9.1.2.3. Reviewer

Members of PACAP routinely review submissions to numerous international conferences and events.

#### 9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

Isabelle Puaut is Associate Editor of IEEE Transactions on Computers (IEEE TC).

André Seznec is a member of the editorial boards of IEEE Micro and ACM Transactions on Architecture and Compiler Optimization.

9.1.3.2. Reviewer - Reviewing Activities

Members of PACAP routinely review submissions to numerous international journals.

#### 9.1.4. Invited Talks

Erven Rohou was invited to the seminar "WCET meets compilation". He presented an invited talk.

#### 9.1.5. Leadership within the Scientific Community

Isabelle Puaut is member of the steering committee of RTNS (Real-Time Networks and Systems).

Isabelle Puaut is member of the steering committee of the Worst Case Execution Time (WCET) workshop, held in conjunction with the Euromicro Conference on Real Time Systems (ECRTS).

#### 9.1.6. Research Administration

Isabelle Puaut is member of the Research Council (Commission Recherche) of the University of Rennes I. She is member of the working group "Habilitation à Diriger des Recherches".

Isabelle Puaut is member of the board of directors (Conseil d'Administration) of ISTIC (computer science and electrical engineering departement of University of Rennes I).

Erven Rohou is a member of the Inria CDT (Commission du Développement Technologique).

As "correspondant scientifique des relations internationales" for Inria Rennes Bretagne Atlantique, Erven Rohou is a member of the Inria COST GTRI (Groupe de Travail "Relations Internationales" du Comité d'Orientation Scientifique et Technologique).

André Seznec is an elected member of the Administration Council of Inria.

#### 9.2. Teaching - Supervision - Juries

#### 9.2.1. Teaching

Licence: Damien Hardy, Linux and C programming, 22 hours, L3, Université de Rennes I, France Licence: Damien Hardy, Real-time systems, 46 hours, L3, Université de Rennes I, France

Master: Damien Hardy, Operating systems, 60 hours, M2, Université de Rennes I, France

Master: Isabelle Puaut, Operating systems: concepts and system programming under Linux (SEL), 75 hours, M1, Université Rennes I, France

Master: Isabelle Puaut, Operating systems internals (NOY), 55 hours, M1, Université Rennes I, France

Master: Isabelle Puaut, Optimizing and Parallelizing Compilers (OPC), 6 hours, M2, Université Rennes I, France

Master: Isabelle Puaut, Real-time systems, 48 hours, M1, Université Rennes I, France

Master: Isabelle Puaut, Writing of scientific publications, 20 hours, M2/PhD, Université Rennes I, France

Master: Sylvain Collange, Parallel Programming, 22 hours, M1, Université Rennes I, France

Master: Sylvain Collange, GPU programming, 32 hours, M2, ESIR, France

Master: Sylvain Collange, Advanced computer architecture, 4 hours, M2, Université Rennes I, France

Master: Sylvain Collange, Advanced CUDA programming, 8 hours, M2, Université Pierre et Marie Curie Paris 6, France

Master: André Seznec, Advanced Architectures, 8 hours, M2, Université de Rennes I, France

#### 9.2.2. Supervision

PhD: Nabil Hallou, *Runtime Optimization of Binary Through Vectorization Transformations*, Université Rennes I, Dec 2017, co-advisors E. Rohou and P. Clauss (EPI Camus Inria Strasbourg)

PhD: Andrea Mondelli, *Revisiting Wide Superscalar Microarchitecture*, Université Rennes I, Sep 2017, co-advisors P. Michaud and A. Seznec

PhD in progress: Rabab Bouziane, *Compilation techniques to exploit novel low-power architecture and technology solutions*, Université Rennes I, started Nov 2015, advisor E. Rohou and Abdoulaye Gamatié (LIRMM, Montpellier)

PhD in progress, Viet Anh Nguyen, *Worst-Case Execution Time (WCET) Estimation for Many-core Architectures*, Université Rennes I, started Jan 2015, co-advisors I. Puaut and D. Hardy.

PhD in progress, Benjamin Rouxel, *Code optimizations for WCET calculation on many-core platforms*, started Oct 2015, co-advisors I. Puaut (70 %) and S. Derrien (30 %) from the CAIRN group.

PhD in progress: Arif Ali Ana-Pparakkal, *Dynamic Function Specialization*, Université Rennes I, started Feb 2015, advisor E. Rohou

PhD in progress: Simon Rokicki, *Compilation dynamique hybride logiciel/matériel*, Université Rennes I, started Sep 2015, co-advisors E. Rohou and Steven Derrien (CAIRN)

PhD in progress: Kalitzidis Kleovoulos, *Ultrawide Issue Superscalar Processors*, Université Rennes I, started Dec 2016, advisor A. Seznec

PhD in progress: Niloofar Charmchi, *Hardware prefetching and related issues*, Université Rennes I, started Jan 2017, advisor A. Seznec

PhD in progress: Daniel Rodriguez Carvalho, *Towards a compressed memory hierarchy*, Université Rennes I, started Oct 2017, advisor A. Seznec

#### 9.2.3. Juries

Isabelle Puaut was a member of the following committees:

- Antoine Blin, Vers une utilisation efficace des processeurs multi-cœurs dans les systèmes embarqués à criticités multiples, Université Pierre et Marie Curie (UPMC), January 2017 (reviewer)
- Laure Abdallah, Worst-case delay analysis of core-to-IO flows over many-cores architectures, Université de Toulouse, April 2017 (reviewer)
- Stefanos Skalistis, Efficient adaptive hard real-time multi-processor systems, École Polytechnique Fédérale de Lausanne (EPFL), October 2017 (reviewer)
- Fabrice Guet, Étude de l'application de la théorie des valeurs extrêmes pour l'estimation fiable et robuste du pire temps d'exécution probabiliste, Université de Toulouse, December 2017 (examiner)
- Nabil Hallou, Runtime optimization of binary through vectorization transformations, Université de Rennes I, December 2017 (examiner)
- Quentin Perret, Predictable execution on many-core processors, Université de Toulouse, April 2017 (examiner)
- Soukayna Msirdi, Modular Avionics Software Integration on Multi-Core COTS, Université de Toulouse, July 2017 (reviewer)

Erven Rohou was a member of the following committees:

- Nabil Hallou, Runtime optimization of binary through vectorization transformations, Université de Rennes I, December 2017.
- Laurent Georget, Suivi de flux d'information correct pour les systèmes d'exploitation Linux, Université de Rennes I, September 2017
- Mohammed Boussaa, Automatic Non-Functional Testing and Tuning of Configurable Generators, Université de Rennes I, September 2017.
- Thierno Barry, Sécurisation à la compilation de logiciels contre les attaques en fautes, Université de Lyon, CEA Grenoble, November 2017.
- Shixiong Xu, Data Layout Oriented Compilation Techniques in Vectorization for Multi-/Manycores, Trinity College, Dublin, Ireland, June 2017.

Erven Rohou was an external expert in the recruitment committee of Alexandra Jimborean, Uppsala Universitet, Sweden.

Erven Rohou was a member of the selection committee of Université de Paris-Est Marne-la-Vallée.

#### 9.3. Popularization

Nicolas Kiss, Damien Hardy and Erven Rohou presented a poster at the "European Cyber Week", organized by the "Pôle d'Excellence Cyber".

Erven Rohou presented a poster at the Teratec Café, describing the ANTAREX H2020 project.

### **10. Bibliography**

#### Major publications by the team in recent years

- [1] F. BODIN, T. KISUKI, P. M. W. KNIJNENBURG, M. F. P. O'BOYLE, E. ROHOU.*Iterative Compilation in a Non-Linear Optimisation Space*, in "Workshop on Profile and Feedback-Directed Compilation (FDO-1), in conjunction with PACT '98", October 1998.
- [2] A. COHEN, E. ROHOU. Processor Virtualization and Split Compilation for Heterogeneous Multicore Embedded Systems, in "DAC", June 2010, p. 102–107.
- [3] N. HALLOU, E. ROHOU, P. CLAUSS, A. KETTERLIN. Dynamic Re-Vectorization of Binary Code, in "SAMOS", July 2015, https://hal.inria.fr/hal-01155207.
- [4] D. HARDY, I. PUAUT. Static probabilistic Worst Case Execution Time Estimation for architectures with Faulty Instruction Caches, in "21st International Conference on Real-Time Networks and Systems", Sophia Antipolis, France, October 2013 [DOI : 10.1145/2516821.2516842], https://hal.inria.fr/hal-00862604.
- [5] D. HARDY, I. SIDERIS, N. LADAS, Y. SAZEIDES. The performance vulnerability of architectural and nonarchitectural arrays to permanent faults, in "MICRO 45", Vancouver, Canada, December 2012, https://hal. inria.fr/hal-00747488.
- [6] D. HARDY, I. SIDERIS, A. SAIDI, Y. SAZEIDES. EETCO: A tool to estimate and explore the implications of datacenter design choices on the tco and the environmental impact, in "Workshop on Energy-efficient Computing for a Sustainable World in conjunction with the 44th Annual IEEE/ACM International Symposium on Microarchitecture (Micro-44)", 2011.
- [7] S. KALATHINGAL, S. COLLANGE, B. NARASIMHA SWAMY, A. SEZNEC. Dynamic Inter-Thread Vectorization Architecture: extracting DLP from TLP, in "International Symposium on Computer Architecture and High-Performance Computing (SBAC-PAD)", Los Angeles, United States, October 2016, https://hal.inria.fr/hal-01356202.
- [8] M.-K. LEE, P. MICHAUD, J. S. SIM, D. NYANG. A simple proof of optimality for the MIN cache replacement policy, in "Information Processing Letters", September 2015, 3 [DOI: 10.1016/J.IPL.2015.09.004], https:// hal.inria.fr/hal-01199424.
- [9] P. MICHAUD. *A Best-Offset Prefetcher Champion*, in "2nd Data Prefetching Championship", Portland, OR, USA, June 2015, https://hal.inria.fr/hal-01165600.

- [10] P. MICHAUD, A. MONDELLI, A. SEZNEC. Revisiting Clustered Microarchitecture for Future Superscalar Cores: A Case for Wide Issue Clusters, in "ACM Transactions on Architecture and Code Optimization (TACO) ", August 2015, vol. 13, n<sup>o</sup> 3, 22 [DOI: 10.1145/2800787], https://hal.inria.fr/hal-01193178.
- [11] P. MICHAUD, A. SEZNEC. Pushing the branch predictability limits with the multi-poTAGE+SC predictor : Champion in the unlimited category, in "4th JILP Workshop on Computer Architecture Competitions (JWAC-4): Championship Branch Prediction (CBP-4)", Minneapolis, United States, June 2014, https://hal.archivesouvertes.fr/hal-01087719.
- [12] A. PERAIS. *Increasing the performance of superscalar processors through value prediction*, Université Rennes 1, September 2015, https://tel.archives-ouvertes.fr/tel-01282474.
- [13] A. PERAIS, A. SEZNEC. EOLE: Paving the Way for an Effective Implementation of Value Prediction, in "International Symposium on Computer Architecture", Minneapolis, MN, United States, ACM/IEEE, June 2014, vol. 42, p. 481 - 492 [DOI: 10.1109/ISCA.2014.6853205], https://hal.inria.fr/hal-01088130.
- [14] A. PERAIS, A. SEZNEC. Practical data value speculation for future high-end processors, in "International Symposium on High Performance Computer Architecture", Orlando, FL, United States, IEEE, February 2014, p. 428 - 439 [DOI: 10.1109/HPCA.2014.6835952], https://hal.inria.fr/hal-01088116.
- [15] A. PERAIS, A. SEZNEC. EOLE: Toward a Practical Implementation of Value Prediction, in "IEEE Micro", June 2015, vol. 35, n<sup>o</sup> 3, p. 114 - 124 [DOI: 10.1109/MM.2015.45], https://hal.inria.fr/hal-01193287.
- [16] E. RIOU, E. ROHOU, P. CLAUSS, N. HALLOU, A. KETTERLIN. *PADRONE: a Platform for Online Profiling, Analysis, and Optimization*, in "Dynamic Compilation Everywhere", Vienna, Austria, January 2014.
- [17] S. SARDASHTI, A. SEZNEC, D. A. WOOD.Skewed Compressed Caches, in "47th Annual IEEE/ACM International Symposium on Microarchitecture, 2014", Minneapolis, United States, December 2014, https:// hal.inria.fr/hal-01088050.
- [18] A. SEMBRANT, T. CARLSON, E. HAGERSTEN, D. BLACK-SHAFFER, A. PERAIS, A. SEZNEC, P. MICHAUD.Long Term Parking (LTP): Criticality-aware Resource Allocation in OOO Processors, in "International Symposium on Microarchitecture, Micro 2015", Honolulu, United States, Proceeding of the International Symposium on Microarchitecture, Micro 2015, ACM, December 2015, https://hal.inria.fr/hal-01225019.
- [19] A. SEZNEC, P. MICHAUD.A case for (partially)-tagged geometric history length predictors, in "Journal of Instruction Level Parallelism", April 2006, http://www.jilp.org/vol8.
- [20] A. SEZNEC, J. SAN MIGUEL, J. ALBERICIO. *The Inner Most Loop Iteration counter: a new dimension in branch history*, in "48th International Symposium On Microarchitecture", Honolulu, United States, ACM, December 2015, 11, https://hal.inria.fr/hal-01208347.
- [21] A. SEZNEC. A New Case for the TAGE Branch Predictor, in "MICRO 2011 : The 44th Annual IEEE/ACM International Symposium on Microarchitecture, 2011", Porto Allegre, Brazil, ACM (editor), ACM-IEEE, December 2011, https://hal.inria.fr/hal-00639193.
- [22] A. SEZNEC. *TAGE-SC-L Branch Predictors: Champion in 32Kbits and 256 Kbits category*, in "JILP Championship Branch Prediction", Minneapolis, United States, June 2014, https://hal.inria.fr/hal-01086920.

- [23] A. SURESH, B. NARASIMHA SWAMY, E. ROHOU, A. SEZNEC. Intercepting Functions for Memoization: A Case Study Using Transcendental Functions, in "ACM Transactions on Architecture and Code Optimization (TACO)", July 2015, vol. 12, n<sup>o</sup> 2, 23 [DOI: 10.1145/2751559], https://hal.inria.fr/hal-01178085.
- [24] A. SURESH. Intercepting functions for memoization, Université Rennes 1, May 2016, https://tel.archivesouvertes.fr/tel-01410539.
- [25] D. D. C. TEIXEIRA, S. COLLANGE, F. M. Q. PEREIRA. Fusion of calling sites, in "International Symposium on Computer Architecture and High-Performance Computing (SBAC-PAD)", Florianópolis, Santa Catarina, Brazil, October 2015 [DOI: 10.1109/SBAC-PAD.2015.16], https://hal.archives-ouvertes.fr/hal-01410221.

#### **Publications of the year**

#### **Doctoral Dissertations and Habilitation Theses**

- [26] N. HALLOU.*Runtime Optimization of Binary Through Vectorization Transformations*, Université de Rennes 1 [UR1], December 2017, https://hal.inria.fr/tel-01672263.
- [27] A. MONDELLI.*Revisiting Wide Superscalar Microarchitecture*, Université de Rennes 1, September 2017, https://hal.inria.fr/tel-01597752.

#### **Articles in International Peer-Reviewed Journal**

- [28] F. ENDO, A. PERAIS, A. SEZNEC. On the Interactions Between Value Prediction and Compiler Optimizations in the Context of EOLE, in "ACM Transactions on Architecture and Code Optimization", June 2017, https:// hal.inria.fr/hal-01519869.
- [29] N. HALLOU, E. ROHOU, P. CLAUSS. Runtime Vectorization Transformations of Binary Code, in "International Journal of Parallel Programming", June 2017, vol. 8, n<sup>o</sup> 6, p. 1536 - 1565 [DOI : 10.1007/s10766-016-0480-z], https://hal.inria.fr/hal-01593216.
- [30] S. KALATHINGAL, S. COLLANGE, B. SWAMY, A. SEZNEC. DITVA: Dynamic Inter-Thread Vectorization Architecture, in "Journal of Parallel and Distributed Computing", 2017, p. 1-32 [DOI: 10.1016/J.JPDC.2017.11.006], https://hal.archives-ouvertes.fr/hal-01655904.
- [31] B. ROUXEL, S. DERRIEN, I. PUAUT. *Tightening Contention Delays While Scheduling Parallel Applications on Multi-core Architectures*, in "ACM Transactions on Embedded Computing Systems (TECS)", October 2017, vol. 16, n<sup>o</sup> 5s, p. 1 20 [DOI: 10.1145/3126496], https://hal.archives-ouvertes.fr/hal-01655383.
- [32] A. SRIDHARAN, B. PANDA, A. SEZNEC. A Band-pass Prefetching : An Effective Prefetch Management Mechanism using Prefetch-fraction Metric in Multi-core Systems, in "ACM Transactions on Architecture and Code Optimization", June 2017, https://hal.inria.fr/hal-01519648.
- [33] A. SRIDHARAN, A. SEZNEC. Dynamic and Discrete Cache Insertion Policies for Managing Shared Last Level Caches in Large Multicores, in "Journal of Parallel and Distributed Computing", May 2017, vol. 106, p. 215–226 [DOI: 10.1016/J.JPDC.2017.02.004], https://hal.inria.fr/hal-01519650.

#### **Invited Conferences**

[34] C. SILVANO, A. BARTOLINI, A. BECCARI, C. MANELFI, C. CAVAZZONI, D. GADIOLI, E. ROHOU, G. PALERMO, G. AGOSTA, J. MARTINOVIČ, J. BISPO, J. M. P. CARDOSO, J. BARBOSA, K. SLANINOVÁ, L. BENINI, M. PALKOVIČ, N. SANNA, P. PINTO, R. CMAR, R. NOBRE, S. CHERUBIN. *The ANTAREX Tool Flow for Monitoring and Autotuning Energy Efficient HPC Systems*, in "SAMOS 2017 - International Conference on Embedded Computer Systems: Architecture, Modeling and Simulation", Pythagorion, Greece, July 2017, https://hal.inria.fr/hal-01615945.

#### **International Conferences with Proceedings**

- [35] V. A. ANH NGUYEN, D. HARDY, I. PUAUT. Cache-conscious offline real-time task scheduling for multicore processors, in "29th Euromicro Conference on Real-Time Systems (ECRTS17)", Dubrovnik, Croatia, Euromicro Conference on Real-Time Systems, June 2017 [DOI: 10.4230/LIPICS.ECRTS.2017.14], http:// hal.upmc.fr/hal-01590421.
- [36] A. A. AP, E. ROHOU. Dynamic Function Specialization, in "International Conference on Embedded Computer Systems: Architectures, MOdeling and Simulation", Pythagorion, Samos, Greece, July 2017, https://hal.inria. fr/hal-01597880.
- [37] R. BOUZIANE, E. ROHOU, A. GAMATIÉ.*How Could Compile-Time Program Analysis help Leveraging Emerging NVM Features?*, in "EDIS 2017 First international conference on Embedded & Distributed Systems", Oran, Algeria, December 2017, p. 1-6, https://hal.inria.fr/hal-01655195.
- [38] R. BOUZIANE, E. ROHOU, A. GAMATIÉ. Compile-Time Silent-Store Elimination for Energy Efficiency: an Analytic Evaluation for Non-Volatile Cache Memory, in "RAPIDO 2018 - 10th Workshop on Rapid Simulation and Performance Evaluation: Methods and Tools", Manchester, United Kingdom, HiPEAC, January 2018, p. 1-8, https://hal.inria.fr/hal-01660686.
- [39] S. CHERUBIN, G. AGOSTA, I. LASRI, E. ROHOU, O. SENTIEYS. *Implications of Reduced-Precision Computations in HPC: Performance, Energy and Error*, in "International Conference on Parallel Computing (ParCo)", Bologna, Italy, September 2017, https://hal.inria.fr/hal-01633790.
- [40] S. COLLANGE.Simty: generalized SIMT execution on RISC-V, in "First Workshop on Computer Architecture Research with RISC-V (CARRV 2017)", Boston, United States, First Workshop on Computer Architecture Research with RISC-V, October 2017, 6, https://hal.inria.fr/hal-01622208.
- [41] S. DERRIEN, I. PUAUT, P. ALEFRAGIS, M. BEDNARA, H. BUCHER, C. DAVID, Y. DEBRAY, U. DURAK, I. FASSI, C. FERDINAND, D. HARDY, A. KRITIKAKOU, G. RAUWERDA, S. REDER, M. SICKS, T. STRIPF, K. SUNESEN, T. TER BRAAK, N. VOROS, J. <sup>†</sup>. BECKER.WCET-aware parallelization of model-based applications for multi-cores: The ARGO approach, in "Design Automation and Test in Europe (DATE), 2017", Lausanne, Switzerland, March 2017, p. 286 289 [DOI: 10.23919/DATE.2017.7927000], http://hal.upmc. fr/hal-01590418.
- [42] D. HARDY, B. ROUXEL, I. PUAUT. *The Heptane Static Worst-Case Execution Time Estimation Tool*, in "17th International Workshop on Worst-Case Execution Time Analysis (WCET 2017)", Dubrovnik, Croatia, International Workshop on Worst-Case Execution Time Analysis, June 2017, vol. 8, 12 [DOI: 10.4230/OASICS.WCET.2017.8], http://hal.upmc.fr/hal-01590444.
- [43] C. MAIZA, P. RAYMOND, C. PARENT-VIGOUROUX, A. BONENFANT, F. CARRIER, H. CASSÉ, P. CUENOT, D. CLARAZ, N. HALBWACHS, E. JAHIER, H. LI, M. DE MICHIEL, V. MUSSOT, I. PUAUT,

C. ROCHANGE, E. ROHOU, J. RUIZ, P. SOTIN, W.-T. SU.*The W-SEPT Project: Towards Semantic-Aware WCET Estimation*, in "17th International Workshop on Worst-Case Execution Time Analysis (WCET 2017)", Dubrovnik, Croatia, International Workshop on Worst-Case Execution Time Analysis, June 2017, 13 [*DOI*: 10.4230/OASICS.WCET.2017.9], http://hal.upmc.fr/hal-01590442.

- [44] S. MARTINEZ, D. HARDY, I. PUAUT. Quantifying WCET reduction of parallel applications by introducing slack time to limit resource contention, in "International Conference on Real-Time Networks and Systems (RTNS), 2017", Grenoble, France, International Conference on Real-Time Networks and Systems, October 2017 [DOI: 10.475/123\_4], http://hal.upmc.fr/hal-01590532.
- [45] R. E. A. MOREIRA, S. COLLANGE, F. M. Q. PEREIRA. Function Call Re-Vectorization, in "ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP)", Austin, Texas, United States, February 2017, https://hal.archives-ouvertes.fr/hal-01410186.
- [46] S. ROKICKI, E. ROHOU, S. DERRIEN. Hardware-Accelerated Dynamic Binary Translation, in "IEEE/ACM Design, Automation & Test in Europe Conference & Exhibition (DATE)", Lausanne, Switzerland, March 2017, https://hal.inria.fr/hal-01423639.
- [47] B. ROUXEL, S. DERRIEN, I. PUAUT. *Tightening contention delays while scheduling parallel applications on multi-core architectures*, in "International Conference on Embedded Software (EMSOFT), 2017", Seoul, South Korea, International Conference on Embedded Software, October 2017, 20 [DOI: 10.1145/3126496], http://hal.upmc.fr/hal-01590508.
- [48] B. ROUXEL, I. PUAUT.STR2RTS: Refactored StreamIT benchmarks into statically analyzable parallel benchmarks for WCET estimation & real-time scheduling, in "17th International Workshop on Worst-Case Execution Time Analysis (WCET 2017)", Dubrovnik, Croatia, June 2017 [DOI : 10.4230/OASICS.WCET.2017.1], http://hal.upmc.fr/hal-01590446.
- [49] M. Y. SIRAICHI, V. F. D. SANTOS, S. COLLANGE, F. M. Q. PEREIRA. *Qubit Allocation*, in "CGO 2018 International Symposium on Code Generation and Optimization", Vienna, Austria, February 2018, p. 1-12 [*DOI* : 10.1145/3168822], https://hal.archives-ouvertes.fr/hal-01655951.
- [50] A. SURESH, E. ROHOU, A. SEZNEC. Compile-Time Function Memoization, in "26th International Conference on Compiler Construction", Austin, United States, February 2017, https://hal.inria.fr/hal-01423811.

#### **National Conferences with Proceeding**

[51] S. COLLANGE, N. BRUNIE. Parcours par liste de chemins : une nouvelle classe de mécanismes de suivi de flot SIMT, in "Conférence d'informatique en Parallélisme, Architecture et Système (ComPAS)", Sophia Antipolis, France, June 2017, https://hal.inria.fr/hal-01522901.

#### **Research Reports**

[52] S. COLLANGE, N. BRUNIE. *Path list traversal: a new class of SIMT flow tracking mechanisms*, Inria Rennes - Bretagne Atlantique, June 2017, n<sup>o</sup> RR-9073, https://hal.inria.fr/hal-01533085.

#### **Other Publications**

[53] S. ROKICKI, E. ROHOU, S. DERRIEN. Supporting Runtime Reconfigurable VLIWs Cores Through Dynamic Binary Translation, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01653110.

#### **References in notes**

- [54] M. HATABA, A. EL-MAHDY, E. ROHOU.OJIT: A Novel Obfuscation Approach Using Standard Just-In-Time Compiler Transformations, in "International Workshop on Dynamic Compilation Everywhere", January 2015.
- [55] R. KUMAR, D. M. TULLSEN, N. P. JOUPPI, P. RANGANATHAN. *Heterogeneous chip multiprocessors*, in "IEEE Computer", nov. 2005, vol. 38, n<sup>0</sup> 11, p. 32–38.
- [56] S. NASSIF, N. MEHTA, Y. CAO.A resilience roadmap, in "Design, Automation Test in Europe Conference Exhibition (DATE), 2010", March 2010, p. 1011-1016.
- [57] J. NICKOLLS, W. J. DALLY. The GPU computing era, in "Micro, IEEE", 2010, vol. 30, nº 2, p. 56-69.
- [58] R. OMAR, A. EL-MAHDY, E. ROHOU. Arbitrary control-flow embedding into multiple threads for obfuscation: a preliminary complexity and performance analysis, in "Proceedings of the 2nd international workshop on Security in cloud computing", ACM, 2014, p. 51–58.
- [59] A. SEZNEC, N. SENDRIER.HAVEGE: A user-level software heuristic for generating empirically strong random numbers, in "ACM Transactions on Modeling and Computer Simulation (TOMACS)", 2003, vol. 13, n<sup>o</sup> 4, p. 334–346.
- [60] H. WONG, T. M. AAMODT. *The Performance Potential for Single Application Heterogeneous Systems*, in "8th Workshop on Duplicating, Deconstructing, and Debunking", 2009.

# **Project-Team PANAMA**

# Parcimonie et Nouveaux Algorithmes pour le Signal et la Modélisation Audio

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: CNRS

Université Rennes 1

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Language, Speech and Audio

### **Table of contents**

1.	Personnel	<b>778</b>
2.	Overall Objectives	779
	2.1. Overall positioning	779
	2.2. Scientific foundations	779
	2.3. Applications	779
3.	Research Program	<b>779</b>
	3.1. Axis 1: Sparse Models and Representations	779
	3.1.1. Efficient Sparse Models and Dictionary Design for Large-scale Data	779
	3.1.2. Compressive Learning	780
	3.2. Axis 2: Robust Acoustic Scene Analysis	780
	3.2.1. Compressive Acquisition and Processing of Acoustic Scenes	780
	3.2.2. Robust Audio Source Separation	780
	3.2.3. Robust Audio Source Localization	781
	3.3. Axis 3: Large-scale Audio Content Processing and Self-organization	781
	3.3.1. Motif Discovery in Audio Data	781
	3.3.2. Structure Modeling and Inference in Audio and Musical Contents	782
4.	Application Domains	782
	4.1. Acoustic Scene Capture	782
	4.2. Audio Signal Separation in Reverberant Environments	782
	4.3. Multimedia Indexing	783
5.	Highlights of the Year	783
6.	New Software and Platforms	
	6.1. VoiceHome Corpus	783
	6.2. FAuST	783
	6.3. SketchMLBox	784
	6.4. SPADE	784
	6.5. FASST	785
	6.6. PHYSALIS	786
7.	New Results	
	7.1. Sparse Representations, Inverse Problems, and Dimension Reduction	786
	7.1.1. Algorithmic and Theoretical results on Computational Representation Learning	786
	7.1.2. Theoretical results on generalized matrix inverses, and the sparse pseudo-inverse	787
	7.1.3. Algorithmic exploration of large-scale Compressive Learning via Sketching	787
	7.1.4. Theoretical results on Low-dimensional Representations, Inverse problems, and D	vimen-
	sion Reduction	788
	7.1.5. Algorithmic Exploration of Sparse Representations in Virtual Reality and Neurofee	dback
		/89
	7.2. Activities on Waveform Design for Telecommunications	/89
	7.3. Emerging activities on Nonlinear Inverse Problems	/90
	7.3.1. Locally-Linear Inverse Regression	/90
	7.3.2. Phase Estimation in Multichannel Mixtures	790
	7.3.3. Audio Inpainting and Denoising	790
	7.4. Source Localization and Separation	/91 701
	7.4.1. Towards Keal-world Separation and Kemixing Applications	791
	7.4.2. Deyond the Local Complex Gaussian Model	792
	7.4.4. Vietually Supervised Auditory Score Archevie	792
	7.4.4. vintually-supervised Auditory Scene Analysis	793
	7.5. INTUSIC Content Processing and Information Retrieval	/94
	1.3.1. Iensor-based Representation of Sectional Units in Music	794

	7.5.2. Modeling music by polytopic graphs of latent relations	794
	7.5.3. Regularity Constraints for the Fusion of Music Structure Segmentation System	795
8.	Bilateral Contracts and Grants with Industry	795
9.	Partnerships and Cooperations	796
	9.1. National Initiatives	796
	9.1.1. Labex Comin Labs projects	796
	9.1.1.1. HEMISFER	796
	9.1.1.2. TEPN	796
	9.1.2. ANR INVATE project with IRT b<>com, Rennes	797
	9.1.3. ANR OATMIL project	797
	9.1.4. OSEO-FUI: voiceHome	798
	9.2. International Initiatives	798
	9.3. International Research Visitors	798
10.	Dissemination	<b>798</b>
	10.1. Promoting Scientific Activities	798
	10.2. Teaching - Supervision - Juries	799
	10.3. Popularization	800
11.	Bibliography	800

### **Project-Team PANAMA**

Creation of the Project-Team: 2013 January 01

#### **Keywords:**

#### **Computer Science and Digital Science:**

- A1.2.6. Sensor networks
- A3.1.1. Modeling, representation
- A3.3.3. Big data analysis
- A3.4.1. Supervised learning
- A3.4.2. Unsupervised learning
- A3.4.4. Optimization and learning
- A3.4.5. Bayesian methods
- A3.4.6. Neural networks
- A3.4.7. Kernel methods
- A3.4.8. Deep learning
- A3.5.1. Analysis of large graphs
- A5.3.2. Sparse modeling and image representation
- A5.7.1. Sound
- A5.7.2. Music
- A5.7.3. Speech
- A5.7.4. Analysis
- A5.9.1. Sampling, acquisition
- A5.9.2. Estimation, modeling
- A5.9.3. Reconstruction, enhancement
- A5.9.4. Signal processing over graphs
- A5.9.5. Sparsity-aware processing
- A5.9.6. Optimization tools
- A5.10.2. Perception
- A5.11.2. Home/building control and interaction
- A6.1.4. Multiscale modeling
- A6.2.5. Numerical Linear Algebra
- A6.2.6. Optimization
- A6.3.1. Inverse problems
- A6.3.2. Data assimilation
- A8.6. Information theory
- A8.7. Graph theory

#### **Other Research Topics and Application Domains:**

- B1.2. Neuroscience and cognitive science
- B2.5.1. Sensorimotor disabilities
- B2.6. Biological and medical imaging
- B5.6. Robotic systems
- B5.8. Learning and training

B6.3.3. - Network Management

- B8.1.2. Sensor networks for smart buildings
- B8.4. Security and personal assistance
- B9.1. Education
- B9.2.1. Music, sound
- B9.2.2. Cinema, Television
- B9.2.3. Video games
- B9.6. Reproducibility
- B9.9.1. Environmental risks

### 1. Personnel

#### **Research Scientists**

Rémi Gribonval [Team leader, Inria, Senior Researcher, HDR] Nancy Bertin [CNRS, Researcher] Frédéric Bimbot [CNRS, Senior Researcher, HDR] Antoine Deleforge [Inria, Researcher] Pierre Vandergheynst [EPFL, Senior Researcher, Chaire Internationale Inria]

#### **Technical Staff**

Nicolas Bellot [Inria, until Apr 2017] Ewen Camberlein [CNRS, until Jul 2017; Inria, since Sep 2017] Romain Lebarbenchon [CNRS, until Jul 2017; Inria, since Sep 2017]

#### **PhD Students**

Antoine Chatalic [Univ de Rennes I, from Sep 2017] Diego Di Carlo [Inria, from Oct 2017] Cássio Fraga Dantas [Inria] Clément Gaultier [Inria] Valentin Gillot [INSA Rennes, from Oct 2017] Mohammed Hafsati [Institut de recherche technologique B-com] Himalaya Jain [Technicolor] Nicolas Keriven [Univ de Rennes I, until Sep 2017] Nathan Libermann [Inria, joint with MULTISPEECH team; in PANAMA from Sep 2017] Corentin Louboutin [Univ de Rennes I]

#### **Post-Doctoral Fellows**

Clément Elvira [Inria, from Dec 2017] Yann Traonmilin [Inria, until Aug 2017]

#### Visiting Scientists

Flavio Castro Alves Teixeira [Univ Insbruck, from May 2017 until Jun 2017] Andreas Loukas [EPFL, from Dec 2017] Helena Peic Tukuljac [EPFL, from Oct 2017] Martin Strauss [Univ Erlangen-Nuremberg, from Oct 2017] Corentin Guichaoua [until Sep 2017]

#### Administrative Assistant

Stéphanie Lemaile [Inria]

### 2. Overall Objectives

#### 2.1. Overall positioning

At the interface between audio modeling and mathematical signal processing, the global objective of PANAMA is to develop mathematically founded and algorithmically efficient techniques to model, acquire and process high-dimensional signals, with a strong emphasis on acoustic data.

Applications fuel the proposed mathematical and statistical frameworks with practical scenarii, and the developed algorithms are extensively tested on targeted applications. PANAMA's methodology relies on a closed loop between theoretical investigations, algorithmic development and empirical studies.

#### 2.2. Scientific foundations

The scientific foundations of PANAMA are focused on sparse representations and probabilistic modeling, and its scientific scope is extended in three major directions:

- The extension of the sparse representation paradigm towards that of "sparse modeling", with the challenge of establishing, strengthening and clarifying connections between sparse representations and machine learning.
- A focus on sophisticated probabilistic models and advanced statistical methods to account for complex dependencies between multi-layered variables (such as in audio-visual streams, musical contents, biomedical data ...).
- The investigation of graph-based representations, processing and transforms, with the goal to describe, model and infer underlying structures within content streams or data sets.

#### 2.3. Applications

The main industrial sectors in relation with the topics of the PANAMA research group are the telecommunication sector, the Internet and multimedia sector, the musical and audiovisual production sector and, marginally, the sector of education and entertainment. Source separation is one of PANAMA's major applicative focus generating increasing industrial transfers. The models, methods and algorithms developed in the team have many potential applications beyond audio processing and modeling – the central theme of the PANAMA project-team – in particular to biomedical signals. Such applications are primarily investigated in partnership with research groups with the relevant expertise (within or outside Inria).

On a regular basis, PANAMA is involved in bilateral or multilateral partnerships, within the framework of consortia, networks, thematic groups, national and European research projects, as well as industrial contracts with various local companies.

### 3. Research Program

#### 3.1. Axis 1: Sparse Models and Representations

#### 3.1.1. Efficient Sparse Models and Dictionary Design for Large-scale Data

Sparse models are at the core of many research domains where the large amount and high-dimensionality of digital data requires concise data descriptions for efficient information processing. Recent breakthroughs have demonstrated the ability of these models to provide concise descriptions of complex data collections, together with algorithms of provable performance and bounded complexity.

A crucial prerequisite for the success of today's methods is the knowledge of a "dictionary" characterizing how to concisely describe the data of interest. Choosing a dictionary is currently something of an "art", relying on expert knowledge and heuristics.

Pre-chosen dictionaries such as wavelets, curvelets or Gabor dictionaries, are based upon stylized signal models and benefit from fast transform algorithms, but they fail to fully describe the content of natural signals and their variability. They do not address the huge diversity underlying modern data much beyond time series and images: data defined on graphs (social networks, internet routing, brain connectivity), vector valued data (diffusion tensor imaging of the brain), multichannel or multi-stream data (audiovisual streams, surveillance networks, multimodal biomedical monitoring).

The alternative to a pre-chosen dictionary is a trained dictionary learned from signal instances. While such representations exhibit good performance on small-scale problems, they are currently limited to low-dimensional signal processing due to the necessary training data, memory requirements and computational complexity. Whether designed or learned from a training corpus, dictionary-based sparse models and the associated methodology fail to scale up to the volume and resolution of modern digital data, for they intrinsically involve difficult linear inverse problems. To overcome this bottleneck, a new generation of efficient sparse models is needed, beyond dictionaries, encompassing the ability to provide sparse and structured data representations as well as computational efficiency. For example, while dictionaries describe low-dimensional signal models in terms of their "synthesis" using few elementary building blocks called atoms, in "analysis" alternatives the low-dimensional structure of the signal is rather "carved out" by a set of equations satisfied by the signal. Linear as well as nonlinear models can be envisioned.

#### 3.1.2. Compressive Learning

A flagship emerging application of sparsity is the paradigm of compressive sensing, which exploits sparse models at the analog and digital levels for the acquisition, compression and transmission of data using limited resources (fewer/less expensive sensors, limited energy consumption and transmission bandwidth, etc.). Besides sparsity, a key pillar of compressive sensing is the use of random low-dimensional projections. Through compressive sensing, random projections have shown their potential to allow drastic dimension reduction with controlled information loss, provided that the projected signal vector admits a sparse representation in some transformed domain. A related scientific domain, where sparsity has been recognized as a key enabling factor, is Machine Learning, where the overall goal is to design statistically founded principles and efficient algorithms in order to infer general properties of large data collections through the observation of a limited number of representative examples. Marrying sparsity and random low-dimensional projections with machine learning shall allow the development of techniques able to efficiently capture and process the information content of large data collections. The expected outcome is a dramatic increase of the impact of sparse models in machine learning, as well as an integrated framework from the signal level (signals and their acquisition) to the semantic level (information and its manipulation), and applications to data sizes and volumes of collections that cannot be handled by current technologies.

#### 3.2. Axis 2: Robust Acoustic Scene Analysis

#### 3.2.1. Compressive Acquisition and Processing of Acoustic Scenes

Acoustic imaging and scene analysis involve acquiring the information content from acoustic fields with a limited number of acoustic sensors. A full 3D+t field at CD quality and Nyquist spatial sampling represents roughly  $10^6$  microphones/ $m^3$ . Dealing with such high-dimensional data requires to drastically reduce the data flow by positioning appropriate sensors, and selecting from all spatial locations the few spots where acoustic sources are active. The main goal is to develop a theoretical and practical understanding of the conditions under which compressive acoustic sensing is both feasible and robust to inaccurate modeling, noisy measures, and partially failing or uncalibrated sensing devices, in various acoustic sensing scenarii. This requires the development of adequate algorithmic tools, numerical simulations, and experimental data in simple settings where hardware prototypes can be implemented.

#### 3.2.2. Robust Audio Source Separation

Audio signal separation consists in extracting the individual sound of different instruments or speakers that were mixed on a recording. It is now successfully addressed in the academic setting of linear instantaneous

mixtures. Yet, real-life recordings, generally associated to reverberant environments, remain an unsolved difficult challenge, especially with many sources and few audio channels. Much of the difficulty comes from the combination of (i) complex source characteristics, (ii) sophisticated underlying mixing model and (iii) adverse recording environments. Moreover, as opposed to the "academic" blind source separation task, most applicative contexts and new interaction paradigms offer a variety of situations in which prior knowledge and adequate interfaces enable the design and the use of informed and/or manually assisted source separation methods.

The former METISS team has developed a generic and flexible probabilistic audio source separation framework that has the ability to combine various acoustic models such as spatial and spectral source models. Building on this existing framework, a first objective of PANAMA is to instantiate and validate specific instances of this framework targeted to real-world industrial applications, such as 5.1 movie re-mastering, interactive music soloist control and outdoor speech enhancement. Extensions of the framework are needed to achieve real-time online processing, and advanced constraints or probabilistic priors for the sources at hand need to be designed, while paying attention to computational scalability issues.

In parallel to these efforts, expected progress in sparse modeling for inverse problems shall bring new approaches to source separation and modeling, as well as to source localization, which is often an important first step in a source separation workflow.

#### 3.2.3. Robust Audio Source Localization

Audio source localization consists in estimating the position of one or several sound sources given the signals received by a microphone array. Knowing the geometry of an audio scene is often a pre-requisite to perform higher-level tasks such as speaker identification and tracking, speech enhancement and recognition or audio source separation. It can be decomposed into two sub-tasks : (i) compute spatial auditory features from raw audio input and (ii) map these features to the desired spatial information. Robustly addressing both these aspects with a limited number of microphones, in the presence of noise, reverberation, multiple and possibly moving sources remains a key challenge in audio signal processing. The first aspect will be tackled by both advanced statistical and acoustical modeling of spatial auditory features. The second one will be addressed by two complementary approaches. *Physics-driven* approaches cast sound source localization as an inverse problem given the known physics of sound propagation within the considered system. Data-driven approaches aim at learning the desired feature-to-source-position mapping using real-world or synthetic training datasets adapted to the problem at hand. Combining these approaches should allow a widening of the notion of source localization, considering problems such as the identification of the directivity or diffuseness of the source as well as some of the boundary conditions of the room. A general perspective is to investigate the relations between the physical structure of the source and the particular structures that can be discovered or enforced in the representations and models used for characterization, localization and separation.

#### 3.3. Axis 3: Large-scale Audio Content Processing and Self-organization

#### 3.3.1. Motif Discovery in Audio Data

Facing the ever-growing quantity of multimedia content, the topic of motif discovery and mining has become an emerging trend in multimedia data processing with the ultimate goal of developing weakly supervised paradigms for content-based analysis and indexing. In this context, speech, audio and music content, offers a particularly relevant information stream from which meaningful information can be extracted to create some form of "audio icons" (key-sounds, jingles, recurrent locutions, musical choruses, etc ...) without resorting to comprehensive inventories of expected patterns.

This challenge raises several fundamental questions that will be among our core preoccupations over the next few years. The first question is the deployment of motif discovery on a large scale, a task that requires extending audio motif discovery approaches to incorporate efficient time series pattern matching methods (fingerprinting, similarity search indexing algorithms, stochastic modeling, etc.). The second question is that of the use and interpretation of the motifs discovered. Linking motif discovery and symbolic learning techniques,

exploiting motif discovery in machine learning are key research directions to enable the interpretation of recurring motifs.

On the application side, several use cases can be envisioned which will benefit from motif discovery deployed on a large scale. For example, in spoken content, word-like repeating fragments can be used for several spoken document-processing tasks such as language-independent topic segmentation or summarization. Recurring motifs can also be used for audio summarization of audio content. More fundamentally, motif discovery paves the way for a shift from supervised learning approaches for content description to unsupervised paradigms where concepts emerge from the data.

#### 3.3.2. Structure Modeling and Inference in Audio and Musical Contents

Structuring information is a key step for the efficient description and learning of all types of contents, and in particular audio and musical contents. Indeed, structure modeling and inference can be understood as the task of detecting dependencies (and thus establishing relationships) between different fragments, parts or sections of information content.

A stake of structure modeling is to enable more robust descriptions of the properties of the content and better model generalization abilities that can be inferred from a particular content, for instance via cache models, trigger models or more general graphical models designed to render the information gained from structural inference. Moreover, the structure itself can become a robust descriptor of the content, which is likely to be more resistant than surface information to a number of operations such as transmission, transduction, copyright infringement or illegal use.

In this context, information theory concepts need to be investigated to provide criteria and paradigms for detecting and modeling structural properties of audio contents, covering potentially a wide range of application domains in speech content mining, music modeling or audio scene monitoring.

### 4. Application Domains

#### 4.1. Acoustic Scene Capture

Acoustic fields carry much information about audio sources (musical instruments, speakers, etc.) and their environment (e.g., church acoustics differ much from office room acoustics). A particular challenge is to capture as much information from a complete 3D+t acoustic field associated with an audio scene, using as few sensors as possible. The feasibility of compressive sensing to address this challenge was shown in certain scenarii, and the actual implementation of this framework will potentially impact practical scenarii such as remote surveillance to detect abnormal events, e.g. for health care of the elderly or public transport surveillance.

#### 4.2. Audio Signal Separation in Reverberant Environments

Audio signal separation consists in extracting the individual sound of different instruments or speakers that were mixed on a recording. It is now successfully addressed in the academic setting of linear instantaneous mixtures. Yet, real-life recordings, generally associated to reverberant environments, remain an unsolved difficult challenge, especially with many sources and few audio channels. Much of the difficulty comes from the estimation of the unknown room impulse response associated to a matrix of mixing filters, which can be expressed as a dictionary-learning problem. Solutions to this problem have the potential to impact, for example, the music and game industry, through the development of new digital re-mastering techniques and virtual reality tools, but also surveillance and monitoring applications, where localizing audio sources is important.

#### 4.3. Multimedia Indexing

Audiovisual and multimedia content generate large data streams (audio, video, associated data such as text, etc.). Manipulating large databases of such content requires efficient techniques to: segment the streams into coherent sequences; label them according to words, language, speaker identity, and more generally to the type of content; index them for easy querying and retrieval, etc. As the next generation of online search engines will need to offer content-based means of searching, the need to drastically reduce the computational burden of these tasks is becoming all the more important as we can envision the end of the era of wasteful datacenters that can increase forever their energy consumption. Most of today's techniques to deal with such large audio streams involve extracting features such as Mel Frequency Cepstral Coefficients (MFCC) and learning high-dimensional statistical models such as Gaussian Mixture Models, with several thousand parameters. The exploration of a compressive learning framework is expected to contribute to new techniques to efficiently process such streams and perform segmentation, classification, etc., in the compressed domain. A particular challenge is to understand how this paradigm can help exploiting truly multimedia features, which combine information from different associated streams such as audio and video, for joint audiovisual processing.

### 5. Highlights of the Year

#### 5.1. Highlights of the Year

#### 5.1.1. Awards

EDF and the French Academy of Technology awarded the 2017 *Paul Caseau Ph.D. prize to Luc Le Magoarou* for his work on "Efficient Matrices for signal processing and machine learning" defended in 2016 [83].

The thesis of Luc Le Magoarou [83] was also awarded a special mention of the AFRIF (the French Association for Shape Recognition and Interpretation) annual Ph.D. prize.

Nicolas Keriven has been awarded the Best Student Paper Award at the international workshop SPARS 2017.

A 2017 EURASIP best paper award was awarded to the paper Universal and efficient compressed sensing by spread spectrum and application to realistic Fourier imaging techniques, co-authored by Gilles Puy, Pierre Vandergheynst, Rémi Gribonval and Yves Wiaux, published in EURASIP Journal on Advances in Signal Processing in 2012 [92].

**BEST PAPERS AWARDS :** 

[] Random Moments for Sketched Mixture Learning.

### 6. New Software and Platforms

#### **6.1. VoiceHome Corpus**

**KEYWORDS:** Audio - Source Separation

FUNCTIONAL DESCRIPTION: This corpus includes reverberated, noisy speech signals spoken by native French talkers in a lounge and recorded by an 8-microphone device at various angles and distances and in various noise conditions. Room impulse responses and noise-only signals recorded in various real rooms and homes and baseline speaker localization and enhancement software are also provided.

- Participants: Ewen Camberlein, Romain Lebarbenchon, Nancy Bertin and Frédéric Bimbot
- Contact: Nancy Bertin
- URL: http://voice-home.gforge.inria.fr/voiceHome\_corpus.html

### 6.2. FAuST

KEYWORDS: Learning - Sparsity - Fast transform - Multilayer sparse factorisation

SCIENTIFIC DESCRIPTION: FAuST allows to approximate a given dense matrix by a product of sparse matrices, with considerable potential gains in terms of storage and speedup for matrix-vector multiplications. FUNCTIONAL DESCRIPTION: Faust 1.x contains Matlab routines to reproduce experiments of the PANAMA team on learned fast transforms.

Faust 2.x contains a C++ implementation with Matlab / Python wrappers (work in progress). NEWS OF THE YEAR: In 2017, new Matlab code for fast approximate Fourier Graph Transforms have been included. based on the approach described in the papers:

-Luc Le Magoarou, Rémi Gribonval, "Are There Approximate Fast Fourier Transforms On Graphs?", ICASSP 2016.

-Luc Le Magoarou, Rémi Gribonval, Nicolas Tremblay, "Approximate fast graph Fourier transforms via multilayer sparse approximations", IEEE Transactions on Signal and Information Processing over Networks, 2017.

- Participants: Luc Le Magoarou, Nicolas Tremblay, Rémi Gribonval, Nicolas Bellot and Adrien Leman
- Contact: Rémi Gribonval
- Publications: Approximate fast graph Fourier transforms via multi-layer sparse approximations - Analyzing the Approximation Error of the Fast Graph Fourier Transform - Flexible Multilayer Sparse Approximations of Matrices and Applications - Are There Approximate Fast Fourier Transforms On Graphs? - Efficient matrices for signal processing and machine learning - FAµST: speeding up linear transforms for tractable inverse problems - Chasing butterflies: In search of efficient dictionaries - Multi-layer Sparse Matrix Factorization
- URL: http://faust.inria.fr/

#### 6.3. SketchMLBox

**KEYWORD:** Clustering

SCIENTIFIC DESCRIPTION: The SketchMLbox is a Matlab toolbox for fitting mixture models to large collections of training vectors using sketching techniques. The collection is first compressed into a vector called sketch, then a mixture model (e.g. a Gaussian Mixture Model) is estimated from this sketch using greedy algorithms typical of sparse recovery. The size of the sketch does not depend on the number of elements in the collection, but rather on the complexity of the problem at hand [2,3]. Its computation can be massively parallelized and distributed over several units. It can also be maintained in an online setting at low cost. Mixtures of Diracs ("K-means") and Gaussian Mixture Models with diagonal covariance are currently available, the toolbox is structured so that new mixture models can be easily implemented

FUNCTIONAL DESCRIPTION: Matlab toolbox for fitting mixture models to large databases using sketching techniques.

- Authors: Nicolas Keriven, Nicolas Tremblay and Rémi Gribonval
- Partner: Université de Rennes 1
- Contact: Rémi Gribonval
- Publications: Sketching for Large-Scale Learning of Mixture Models Compressive K-means

   Spikes super-resolution with random Fourier sampling Sketching for large-scale learning of
   mixture models Blind Source Separation Using Mixtures of Alpha-Stable Distributions Sketching
   for Large-Scale Learning of Mixture Models Compressive Gaussian Mixture Estimation by
   Orthogonal Matching Pursuit with Replacement
- URL: http://sketchml.gforge.inria.fr

#### 6.4. SPADE

Sparse Audio Declipper KEYWORDS: Audio - Sparse regularization - Declipping SCIENTIFIC DESCRIPTION: SPADE (the Sparse Audio Declipper) allows to reproduce audio declipping experiments from the papers:

- Srdan Kitic, Nancy Bertin, Remi Gribonval. Audio Declipping by Cosparse Hard Thresholding. iTwist - 2nd international - Traveling Workshop on Interactions between Sparse models and Technology, Aug 2014, Namur, Belgium.

- Srdan Kitic, Nancy Bertin, Remi Gribonval. Sparsity and cosparsity for audio declipping: a flexible nonconvex approach. LVA/ICA 2015 - The 12th International Conference on Latent Variable Analysis and Signal Separation, Aug 2015, Liberec, Czech Republic.

FUNCTIONAL DESCRIPTION: SPADE is a declipping algorithm developed by the PANAMA project-team. To the best of our knowledge SPADE achieves state-of-the-art audio declipping quality. Real-time processing of audio streams is possible.

The web site http://spade.inria.fr provides example audio files and allows users to test SPADE on their own files, either by downloading Matlab routines or using Inria's software demonstration platform, Allgo, to test it on the web.

NEWS OF THE YEAR: In 2017, a web interface to demonstrate the potential of SPADE has been setup using the Allgo platform.

- Participants: Nancy Bertin, Clement Gaultier, Ewen Camberlein, Romain Lebarbenchon, Rémi Gribonval and Srdan Kitic
- Contact: Rémi Gribonval
- Publications: Audio Declipping by Cosparse Hard Thresholding Sparsity and cosparsity for audio declipping: a flexible non-convex approach
- URL: http://spade.inria.fr/

#### 6.5. FASST

Flexible Audio Source Separation Toolbox

KEYWORD: Audio signal processing

SCIENTIFIC DESCRIPTION: FASST is a Flexible Audio Source Separation Toolbox, designed to speed up the conception and automate the implementation of new model-based audio source separation algorithms.

FASST 1.0 development was achieved by the METISS team in Rennes and is now deprecated.

FASST 2.1 (current version) development was jointly achieved by the PAROLE team in Nancy and the (former) TEXMEX team in Rennes through an Inria funded ADT (Action de Développement Technologique). PANAMA contributed to the development by coordinating and performing user tests, and to the dissemination in a Show-and-Tell ICASSP poster [58]. While the first implementation was in Matlab, the new implementation is in C++ (for core functions), with Matlab and Python user scripts. Version 2, including speedup and new features was released in 2014 and can be downloaded from http://bass-db.gforge.inria.fr/fasst/.

A new version is currently under development in the PANAMA team through the Inria funded ADT "FFWD" (FASST For Wider Dissemination) and will be released in 2018.

FUNCTIONAL DESCRIPTION: FASST is a Flexible Audio Source Separation Toolbox designed to speed up the conception and automate the implementation of new model-based audio source separation algorithms. It is the only audio source separation software available to the public (QPL licence) which simultaneously exploits spatial and spectral cues on the sources to separate.

- Participants: Alexey Ozerov, Nancy Bertin, Ewen Camberlein, Romain Lebarbenchon, Emmanuel Vincent, Frédéric Bimbot and Yann Salaun
- Contact: Emmanuel Vincent
- URL: http://bass-db.gforge.inria.fr/fasst/

### 6.6. PHYSALIS

**KEYWORDS: Source localization - Cosparsity** 

SCIENTIFIC DESCRIPTION: PHYSALIS (Physics-Driven Cosparse Analysis) gathers algorithms for (joint) source localization and estimation, expressed as inverse problems and addressed with co-sparse regularization. A particular emphasis is put on the acoustic and EEG settings.

FUNCTIONAL DESCRIPTION: PHYSALIS is distributed as a set of Matlab routines to reproduce experimental results from the Ph.D. thesis of Srdan Kitic.

NEWS OF THE YEAR: In 2017, the code of PHYSALIS has been packaged at the occasion of the writing of an overview chapter on co-sparse source localization.

- Participants: Laurent Albera, Nancy Bertin, Rémi Gribonval and Srdan Kitic
- Contact: Rémi Gribonval
- Publications: Physics-driven inverse problems made tractable with cosparse regularization -Cosparse regularization of physics-driven inverse problems - Versatile and scalable cosparse methods for physics-driven inverse problems - Hearing behind walls: localizing sources in the room next door with cosparsity - Sparse Acoustic Source Localization with Blind Calibration for Unknown Medium Characteristics - The best of both worlds: synthesis-based acceleration for physics-driven cosparse regularization
- URL: http://cosoloc.gforge.inria.fr/

### 7. New Results

#### 7.1. Sparse Representations, Inverse Problems, and Dimension Reduction

Sparsity, low-rank, dimension-reduction, inverse problem, sparse recovery, scalability, compressive sensing The team has had a substantial activity ranging from theoretical results to algorithmic design and software contributions in the fields of sparse representations, inverse problems, and dimension reduction.

#### 7.1.1. Algorithmic and Theoretical results on Computational Representation Learning

Participants: Rémi Gribonval, Nicolas Bellot, Cássio Fraga Dantas.

Main collaborations: Luc Le Magoarou (IRT b<>com, Rennes), Nicolas Tremblay (GIPSA-Lab, Grenoble), R. R. Lopes and M. N. Da Costa (DSPCom, Univ. Campinas, Brazil)

An important practical problem in sparse modeling is to choose the adequate dictionary to model a class of signals or images of interest. While diverse heuristic techniques have been proposed in the literature to learn a dictionary from a collection of training samples, classical dictionary learning is limited to small-scale problems. Inspired by usual fast transforms, we proposed a general dictionary structure (called FA $\mu$ ST for Flexible Approximate Multilayer Sparse Transforms) that allows cheaper manipulation, and an algorithm to learn such dictionaries together with their fast implementation.

The principle and its application to image denoising appeared at ICASSP 2015 [80] and an application to speedup linear inverse problems was published at EUSIPCO 2015 [79]. A Matlab library has been released (see FA $\mu$ ST in Section 6.2) to reproduce the experiments from the comprehensive journal paper published in 2016 [82], which additionally includes theoretical results on the improved sample complexity of learning such dictionaries. Pioneering identifiability results have been obtained in the Ph.D. thesis of Luc Le Magoarou on this topic [83].

We further explored the application of this technique to obtain fast approximations of Graph Fourier Transforms. A conference paper on this latter topic appeared in ICASSP 2016 [81], and a journal paper has been published this year [17] where we empirically show that  $O(n \log n)$  approximate implementations of Graph Fourier Transforms are possible for certain families of graphs. This opens the way to substantial accelerations for Fourier Transforms on large graphs. The approximation error of such Fast Graph Fourier Transforms has been studied in a conference paper [31].

A C++ version of the FA $\mu$ ST software library has been developed (see Section 6) to release the resulting algorithms and interface them with both Matlab and Python (work in progress).

As a complement to the FA $\mu$ ST structure for matrix approximation, we proposed a learning algorithm that constrains the dictionary to be a sum of Kronecker products of smaller sub-dictionaries. A special case of the proposed structure is the widespread separable dictionary. This approach, named SuKro, was evaluated experimentally on an image denoising application [39].

We combined accelerated matrix-vector multiplications offered by FAuST matrix approximations with dynamic screening [52], that safely eliminates inactive variables to speedup iterative sparse recovery algorithms. First, we showed how to obtain safe screening rules for the exact problem while manipulating an approximate dictionary. We then adapted an existing screening rule to this new framework and define a general procedure to leverage the advantages of both strategies. Significant complexity reductions were obtained in comparison to screening rules alone [35].

#### 7.1.2. Theoretical results on generalized matrix inverses, and the sparse pseudo-inverse Participant: Rémi Gribonval.

#### Main collaboration: Ivan Dokmanic (University of Illinois at Urbana Champaign, USA)

We studied linear generalized inverses that minimize matrix norms. Such generalized inverses are famously represented by the Moore-Penrose pseudoinverse (MPP) which happens to minimize the Frobenius norm. Freeing up the degrees of freedom associated with Frobenius optimality enables us to promote other interesting properties. In a first part of this work [37], we looked at the basic properties of norm-minimizing generalized inverses, especially in terms of uniqueness and relation to the MPP. We first showed that the MPP minimizes many norms beyond those unitarily invariant, thus further bolstering its role as a robust choice in many situations. We then concentrated on some norms which are generally not minimized by the MPP, but whose minimization is relevant for linear inverse problems and sparse representations. In particular, we looked at mixed norms and the induced  $\ell^p \rightarrow \ell^q$  norms.

An interesting representative is the sparse pseudoinverse which we studied in much more detail in a second part of this work [38], motivated by the idea to replace the Moore-Penrose pseudoinverse by a sparser generalized inverse which is in some sense well-behaved. Sparsity implies that it is faster to apply the resulting matrix; well-behavedness would imply that we do not lose much in stability with respect to the least-squares performance of the MPP. We first addressed questions of uniqueness and non-zero count of (putative) sparse pseudoinverses. We showed that a sparse pseudoinverse is generically unique, and that it indeed reaches optimal sparsity for almost all matrices. We then turned to proving a stability result: finite-size concentration bounds for the Frobenius norm of p-minimal inverses for  $1 \le p \le 2$ . Our proof is based on tools from convex analysis and random matrix theory, in particular the recently developed convex Gaussian min-max theorem. Along the way we proved several results about sparse representations and convex programming that were known folklore, but of which we could find no proof.

#### 7.1.3. Algorithmic exploration of large-scale Compressive Learning via Sketching

Participants: Rémi Gribonval, Nicolas Keriven, Antoine Chatalic, Antoine Deleforge.

Main collaborations: Patrick Perez (Technicolor R&I France, Rennes), Anthony Bourrier (formerly Technicolor R&I France, Rennes; then GIPSA-Lab, Grenoble), Antoine Liutkus (ZENITH Inria project-team, Montpellier), Nicolas Tremblay (GIPSA-Lab, Grenoble), Phil Schniter & Evan Byrne (Ohio State University, USA)

**Sketching for Large-Scale Mixture Estimation.** When fitting a probability model to voluminous data, memory and computational time can become prohibitive. We proposed during the Ph.D. thesis of Anthony Bourrier [53], [56], [54], [55] a framework aimed at fitting a mixture of isotropic Gaussians to data vectors by computing a low-dimensional sketch of the data. The sketch represents empirical moments of the underlying probability distribution. Deriving a reconstruction algorithm by analogy with compressive sensing, we experimentally showed that it is possible to precisely estimate the mixture parameters provided that the sketch is large enough. The proposed algorithm provided good reconstruction and scaled to higher dimensions than previous probability mixture estimation algorithms, while consuming less memory in the case of voluminous datasets. It

also provided a potentially privacy-preserving data analysis tool, since the sketch does not explicitly disclose information about individual datum.

During the Ph.D. thesis of Nicolas Keriven [12], we consolidated our extensions to non-isotropic Gaussians, with a new algorithm called CL-OMP [72] and conducted large-scale experiments demonstrating its potential for speaker verification. A conference paper appeared at ICASSP 2016 [71] and the journal version has been accepted this year [44], accompanied by a toolbox for reproducible research (see SketchMLBox, Section 6.3). Nicolas Keriven was awarded the SPARS 2017 Best Student Paper Award for this work [].

**Sketching for Compressive Clustering and beyond.** Last year we started a new endeavor to extend the approach beyond the case of Gaussian Mixture Estimation.

First, we showed empirically that sketching can be adapted to compress a training collection while still allowing large-scale *clustering*. The approach, called "Compressive K-means", uses CL-OMP at the learning stage and is described in a paper published at ICASSP 2017 [23]. In the high-dimensional setting, it is also possible to substantially speedup both the sketching stage and the learning stage by replacing Gaussian random matrices with fast random matrices in the sketching procedure. This has been demonstrated by Antoine Chatalic during his internship and submitted for publication to a conference. An alternative algorithm for cluster recovery from a sketch was proposed this year, based on simplified hybrid generalized approximate message passing (SHyGAMP). Numerical experiments suggest that this approach is more efficient than CL-OMP (in both computational and sample complexity) and more efficient than k-means++ in certain regimes [25].

Then, we leveraged the parallel between the mathematical expression of sketched clustering and superresolution to explore the potential of sketching and CL-OMP for the stable recovery of signals made of few spikes (in the gridless setting) from few random weighted Fourier measurements [47]. We also demonstrated that sketching can be used in blind source localization and separation, by learning mixtures of alpha-stable distributions [45], see details in Section 7.4.2.

# 7.1.4. Theoretical results on Low-dimensional Representations, Inverse problems, and Dimension Reduction

Participants: Rémi Gribonval, Yann Traonmilin.

Main collaboration: Mike Davies (University of Edinburgh, UK), Gilles Puy (Technicolor R&I France, Rennes),

#### Inverse problems and compressive sensing in Hilbert spaces.

Many inverse problems in signal processing deal with the robust estimation of unknown data from underdetermined linear observations. Low dimensional models, when combined with appropriate regularizers, have been shown to be efficient at performing this task. Sparse models with the  $\ell^1$ -norm or low-rank models with the nuclear norm are examples of such successful combinations. Stable recovery guarantees in these settings have been established using a common tool adapted to each case: the notion of restricted isometry property (RIP). Last year we published a comprehensive paper [96] establishing generic RIP-based guarantees for the stable recovery of cones (positively homogeneous model sets) with arbitrary regularizers. We also described a generic technique to construct linear maps from a Hilbert space to  $\mathbb{R}^m$  that satisfy the RIP [20]. These results have been surveyed in a book chapter completed this year[48].

**Information preservation guarantees with low-dimensional sketches.** We established a theoretical framework for sketched learning, encompassing statistical learning guarantees as well as dimension reduction guarantees. The framework provides theoretical grounds supporting the experimental success of our algorithmic approaches to compressive K-means, compressive Gaussian Mixture Modeling, as well as compressive Principal Component Analysis (PCA). A comprehensive preprint has been completed and submitted to a journal [42]. Future work will include expliciting the impact of the proposed framework on a wider set of concrete learning problems.

# 7.1.5. Algorithmic Exploration of Sparse Representations in Virtual Reality and Neurofeedback

Participant: Rémi Gribonval.

Ferran Argelaget & Anatole Lecuyer (HYBRID Inria project-team, Rennes), Saman Noorzadeh, Pierre Maurel & Christian Barillot (VISAGES Inria project-team, Rennes)

In collaboration with the VISAGES team we validated a technique to estimate brain neuronal activity by combining EEG and fMRI modalities in a joint framework exploiting sparsity [34].

Our work in collaboration with the HYBRID team on sparse dictionary learning for spatial and rotation invariant gesture recognition has been published this year [24]. Our work on multi-modal

7.1.5.1. An Alternative Framework for Sparse Representations: Sparse "Analysis" Models

Participants: Rémi Gribonval, Nancy Bertin, Clément Gaultier.

Main collaborations: Srdan Kitic (Technicolor R & I France, Rennes), Laurent Albera and Siouar Bensaid (LTSI, Univ. Rennes)

In the past decade there has been a great interest in a synthesis-based model for signals, based on sparse and redundant representations. Such a model assumes that the signal of interest can be composed as a linear combination of *few* columns from a given matrix (the dictionary). An alternative *analysis-based* model can be envisioned, where an analysis operator multiplies the signal, leading to a *cosparse* outcome.

Building on our pioneering work on the cosparse model [70], [89][8], successful applications of this approach to sound source localization, brain imaging and audio restoration have been developed in the team during the last years [73], [75], [74], [50]. Along this line, two main achievements were obtained this year. First, and following the publication in 2016 of a journal paper embedding in a unified fashion our results in source localization [5], we wrote a book chapter (currently in press) gathering our contributions in physics-driven cosparse regularization, including new results and algorithms demonstrating the versatility, robustness and computational efficiency of our methods in realistic, large scale scenarios in acoustics and EEG signal processing [46]. Second, we continued extending the cosparse framework on audio restoration problems: improvements on our released real-time declipping algorithm (A-SPADE - see Section 6), new results on the denoising task [41], [28], and the submission of a journal paper encompassing several denoising and declipping methods in a common framework [40].

#### 7.2. Activities on Waveform Design for Telecommunications

Peak to Average Power Ratio (PAPR), Orthogonal Frequency Division Multiplexing (OFDM), Generalized Waveforms for Multi Carrier (GWMC), Adaptive Wavelet Packet Modulation (AWPM)

#### 7.2.1. Characterizing and designing multi-carrier waveform systems with optimum PAPR Participant: Rémi Gribonval.

#### Main collaboration: Marwa Chafii, Jacques Palicot, Carlos Bader (SCEE team, CentraleSupelec, Rennes)

In the context of the TEPN (Towards Energy Proportional Networks) Comin Labs project (see Section 9.1.1.2), in collaboration with the SCEE team at Supelec (thesis of Marwa Chafii [57], defended in October 2016 and co-supervised by R. Gribonval), we investigated a problem related to dictionary design: the characterization of waveforms with low Peak to Average Power Ratio (PAPR) for wireless communications. This is motivated by the importance of a low PAPR for energy-efficient transmission systems.

A first stage of the work consisted in characterizing the statistical distribution of the PAPR for a general family of multi-carrier systems, leading to a journal paper [61] and several conference communications [59], [60]. Our characterization of waveforms with optimum PAPR [62] has been published in a journal in 2016 [58]. Our work on the design of new adaptive multi-carrier waveform systems able to cope with frequency-selective channels while minimizing PAPR which gave rise to a patent in 2016 [63] has been been submitted for publication as a journal paper. Our study of the tradeoffs between PAPR and Power Spectral Density properties of a wavelet modulation scheme has been published this year [14].

#### 7.3. Emerging activities on Nonlinear Inverse Problems

Compressive sensing, compressive learning, audio inpainting, phase estimation

#### 7.3.1. Locally-Linear Inverse Regression

Participant: Antoine Deleforge.

Main collaborations: Florence Forbes (MISTIS Inria project-team, Grenoble), Emeline Perthame (HUB team, Institut Pasteur, Paris), Vincent Drouard, Radu Horaud, Sileye Ba and Georgios Evangelidis (PERCEPTION Inria project-team, Grenoble)

A general problem in machine learning and statistics is that of *high- to low-dimensional mapping*. In other words, given two spaces  $\mathbb{R}^D$  and  $\mathbb{R}^L$  with  $D \gg L$ , how to find a relation between these two spaces such that given a new observation vector  $y \in \mathbb{R}^D$  its associated vector  $x \in \mathbb{R}^L$  can be estimated? In *regression*, a set of training pairs  $\{(y_n, x_n)\}_{n=1}^N$  is used to learn the relation. In *dimensionality reduction*, only vectors  $\{y_n\}_{n=1}^N$  are observed, and an intrinsic low-dimensional representation  $\{x_n\}_{n=1}^N$  is sought. In [66], we introduced a probabilistic framework unifying both tasks referred to as *Gaussian Locally Linear Mapping* (GLLiM). The key idea is to learn an easier other-way-around locally-linear relationship from x to y using a joint Gaussian Mixture model on x and y. This mapping is then easily reversed via Bayes' inversion. This framework was notably applied to hyperspectral imaging of Mars [64], head pose estimation in images [16], sound source separation and localization [65], and virtually-supervised acoustic space learning (see Section 7.4.4). This year, in [19], we introduced the *Student Locally Linear Mapping* (SLLiM) framework. The use of heavy-tailed Student's t-distributions instead of Gaussian ones leads to more robustness and better regression performance on several datasets.

#### 7.3.2. Phase Estimation in Multichannel Mixtures

Participants: Antoine Deleforge, Yann Traonmilin.

Main collaboration: Angélique Drémeau (ENSTA Bretagne and Lab-STICC, Brest)

The problem of estimating source signals given an observed multichannel mixture is fundamentally ill-posed when the mixing matrix is unknown or when the number of sources is larger that the number of microphones. Hence, prior information on the desired source signals must be incorporated in order to tackle it. An important line of research in audio source separation over the past decade consists in using a model of the source signals' magnitudes in the short-time Fourier domain [9]. Such models can be inferred through, e.g., nonnegative matrix factorization [9] or deep neural networks [90]. Magnitudes estimates are often interpreted as instantaneous variances of Gaussian-process source signals, and are combined with Wiener filtering for source separation. In [26], we introduced a shift of this paradigm by considering the *Phase Unmixing* problem: how can one recover the instantaneous phases of complex mixed source signals when their magnitudes and mixing matrix are known? This problem was showed to be NP-hard, and three approaches were proposed to tackle it: a heuristic method, an alternate minimization method, and a convex relaxation into a semi-definite program. The last two approaches were showed to outperform the oracle multichannel Wiener filter in underdetermined informed source separation tasks. The latter yielded best results, including the potential for exact source separation in under-determined settings. In [27] we applied this framework to the classical problem of *phase retrieval* with a novel multivariate Von Mises prior on phases. We showed that enforcing this prior yielded more accurate estimates than state-of-the art phase retrieval methods.

#### 7.3.3. Audio Inpainting and Denoising

Participants: Rémi Gribonval, Nancy Bertin, Clément Gaultier.

Main collaborations: Srdan Kitic (Technicolor R&I France, Rennes)

Inpainting is a particular kind of inverse problems that has been extensively addressed in the recent years in the field of image processing. Building upon our previous pioneering contributions (definition of the audio inpainting problem as a general framework for many audio processing tasks, application to the audio declipping or desaturation problem, formulation as a sparse recovery problem [49]), we proposed over the last two years a series of algorithms leveraging the competitive cosparse approach, which offers a very appealing trade-off between reconstruction performance and computational time [74], [77] [6]. The work on cosparse audio declipping which was awarded the Conexant best paper award at the LVA/ICA 2015 conference [77]

In 2017, this work was extended towards advanced (co)sparse decompositions, including several forms of structured sparsity in the time-frequency domain and across channels, and towards their application to the denoising task, in addition to the previously introduced declipping task, which we continued to improve. In particular, we investigated the incorporation of the so-called "social" structure constraint [78] into problems regularized by a cosparse prior [28], [41], and exhibited a common framework allowing to tackle both denoising and declipping in a unified fashion [40]. A new algorithm for joint declipping of multichannel audio was also derived (one submitted conference publication.)

#### 7.4. Source Localization and Separation

resulted in a software release in 2016.

Source separation, sparse representations, probabilistic model, source localization

Acoustic source localization is, in general, the problem of determining the spatial coordinates of one or several sound sources based on microphone recordings. This problem arises in many different fields (speech and sound enhancement, speech recognition, acoustic tomography, robotics, aeroacoustics...) and its resolution, beyond an interest in itself, can also be the key preamble to efficient source separation, which is the task of retrieving the source signals underlying a multichannel mixture signal.

Over the last years, we proposed a general probabilistic framework for the joint exploitation of spatial and spectral cues [9], hereafter summarized as the "local Gaussian modeling", and we showed how it could be used to quickly design new models adapted to the data at hand and estimate its parameters via the EM algorithm. This model became the basis of a large number of works in the field, including our own. This accumulated progress lead, in 2015, to two main achievements: a new version of the Flexible Audio Source Separation Toolbox, fully reimplemented, was released [93] and we published an overview paper on recent and going research along the path of *guided* separation in a special issue of IEEE Signal Processing Magazine [10].

From there, our recent work divided into several tracks: maturity work on the concrete use of these tools and principles in real-world scenarios, in particular within the voiceHome and INVATE projects (see Section 7.4.1); more exploratory work towards new approaches diverging away from local Gaussian modeling (Section 7.4.2); formulating and addressing a larger class of problems related to localization and separation, in the context of robotics (Section 7.4.3) and audio scene analysis with machine learning (Section 7.4.4).

#### 7.4.1. Towards Real-world Separation and Remixing Applications

**Participants:** Nancy Bertin, Frédéric Bimbot, Rémi Gribonval, Ewen Camberlein, Romain Lebarbenchon, Mohammed Hafsati.

## Main collaborations: Emmanuel Vincent (MULTISPEECH Inria project-team, Nancy), Nicolas Epain (IRT b<>com, Rennes)

Based on the team's accumulated expertise and tools for localization and separation using the local Gaussian model, two real-world applications were addressed in the past year, which in turn gave rise to new research tracks.

First, we were part of the voiceHome project (2015-2017, see Section 9.1.4), an industrial collaboration aiming at developing natural language dialog in home applications, such as control of domotic and multimedia devices, in realistic and challenging situations (very noisy and reverberant environments, distant microphones). We benchmarked, improved and optimized existing localization and separation tools to the particular context of this application, worked on a better interface between source localization and source separations steps

and on optimal initialization scenarios, and reduced the latency and computational burden of the previously available tools, highlighting operating conditions were real-time processing is achievable. Automatic selection of the best microphones subset in an array was investigated. A journal publication including new data (extending the voiceHome Corpus, see Section 6.1), baseline tools and results was submitted to a special issue of Speech Communication. Accomplished progress and levers of improvements identified thanks to this project resulted in the granting of an Inria ADT (Action de Développement Technologique), which started in September 2017, for a new development phase of the FASST software (see Section 6.5).

Second, through the Ph.D. of Mohammed Hafsati (in collaboration with the IRT b <> com with the INVATE project, see Section 9.1.2) started in November 2016, we investigated a new application of source separation to sound re-spatialization from Higher Order Ambisonics (HOA) signals [69], in the context of free navigation in 3D audiovisual contents. We studied the applicability conditions of the FASST framework to HOA signals and benchmarked localization and separation methods in this domain. We started extending our methods to hybrid acquisition scenarios, where the separation of HOA signals can be informed by the complementary close-up microphonic signals. Future work will include systematic experimental evaluation.

#### 7.4.2. Beyond the Local Complex Gaussian Model

Participants: Antoine Deleforge, Nicolas Keriven.

#### Main collaboration: Antoine Liutkus (ZENITH Inria project-team, Montpellier)

The team has also recently investigated a number of alternative probabilistic models to the local complex Gaussian (LCG) model for audio source separation. An important limit of LCG is that most signals of interest such as speech or music do not exhibit Gaussian distributions but heavier-tailed ones due to their important dynamic [84]. In [45] we proposed a new sound source separation algorithm using heavy-tailed alpha stable priors for source signals. Experiments showed that it outperformed baseline Gaussian-based methods on underdetermined speech or music mixtures. Another limitation of LCG is that it implies a zero-mean complex prior on source signals. This induces a bias towards low signal energies, in particular in under-determined settings. With the development of accurate magnitude spectrogram models for audio signals such as nonnegative matrix factorization [91][9] or more recently deep neural networks [90], it becomes desirable to use probabilistic models enforcing strong magnitude priors. In [26], we explored deterministic magnitude models (see section 7.3.2 for details). An approximate and tractable probabilistic version of this referred to as BEADS (Bayesian Expansion Approximating the Donut Shape) is currently under development. The source prior considered is a mixture of isotropic Gaussians regularly placed on a zero-centered complex circle.

#### 7.4.3. Applications to Robot Audition

Participants: Nancy Bertin, Antoine Deleforge, Martin Strauss, Victor Miguet.

## Main collaborations: Aly Magassouba, Pol Mordel and François Chaumette (LAGADIC Inria project-team, Rennes), Alexander Schmidt and Walter Kellermann (University of Erlangen-Nuremberg, Germany)

**Implicit Localization through Audio-based Control.** In robotics, the use of aural perception has received recently a growing interest but still remains marginal in comparison to vision. Yet audio sensing is a valid alternative or complement to vision in robotics, for instance in homing tasks. Most existing works are based on the relative localization of a defined system with respect to a sound source, and the control scheme is generally designed separately from the localization system. In contrast, the approach that we investigated in the context of Aly Magassouba's Ph.D. (defended in December 2016) focused on a sensor-based control approach. A journal paper encompassing and extending the results obtained before 2017 [88], [86], [87] has been submitted to IEEE Transactions on Robotics (accepted with minor revisions). In 2017, we obtained new results on the use of interaural level difference as the only input feature of the servo, with new experimental validation on humanoid robots. A publication about these last results has been submitted to IEEE Robotics and Automation Letters.
**Ego-noise Reduction with Motor-Data-Guided Dictionary Learning.** Ego-noise reduction is the problem of suppressing the noise a robot caused by its own motions. Such noise degrades the recorded microphone signal such that the robot's auditory capabilities suffer. To suppress it, it is intuitive to use also motor data, since it provides additional information about the robot's joints and thereby the noise sources. In [95], we incorporated motor data to a recently proposed multichannel dictionary algorithm [68]. We applied this to ego-noise reduction on the humanoid robot NAO. At training, a dictionary is learned that captures spatial and spectral characteristics of ego-noise. At testing, nonlinear classifiers are used to efficiently associate the current robot's motor state to relevant sets of entries in the learned dictionary. By this, computational load is reduced by one third in typical scenarios while achieving at least the same noise reduction performance. Moreover, we proposed to train dictionaries on different microphone array geometries and used them for ego-noise reduction while the head on which the microphones are mounted is moving. In such scenarios, the motor-data-guided approach resulted in significantly better performance values.

Sound Source Localization with a Drone. Flying robots or drones have undergone a massive development in recent years. Already broadly commercialized for entertainment purpose, they also underpin a number of exciting future applications such as mail delivery, smart agriculture, archaeology or search and rescue. An important technological challenge for these platforms is that of localizing sound sources in order to better analyse and understand their environment. For instance, how to localize a person crying for help in the context of a natural disaster? This challenge raises a number of difficult scientific questions. How to efficiently embed a microphone array on a drone? How to deal with the heavy ego-noise produced by the drone's motors? How to deal with moving microphones and distant sources? Victor Miguet and Martin Strauss tackled part of these challenges during their masters' internships. A light 3D-printed structure was designed to embed a USB sound card and a cubic 8-microphone array under a Mikrokopter drone that can carry up to 800 g of payload in flights. Noiseless speech and on-flights ego-noise datasets were recorded. The data were precisely annotated with the target source's position, the state of each drone's propellers and the drone's position and velocity. Baseline methods including multichannel Wiener filtering, GCC-PHAT and MUSIC were implemented in both C++ and Matlab and were tested on the dataset. Up to  $5^{\circ}$  speech localization accuracy in both azimuth and elevation was achieved under heavy-noise conditions (-5 dB signal-to-noise-ratio). We plan to make the datasets and code publicly available in 2018.

#### 7.4.4. Virtually-Supervised Auditory Scene Analysis

Participants: Antoine Deleforge, Nancy Bertin, Diego Di Carlo, Clément Gaultier.

Main collaborations: Ivan Dokmanic (University of Illinois at Urbana-Champaign, Coordinated Science Lab, USA) and Robin Scheibler (Tokyo Metropolitan University, Tokyo, Japan), Saurabh Kataria (IIT Kanpur, India)

Classical audio signal processing methods strongly rely on a good knowledge of the *geometry* of the audio scene, *i.e.*, what are the positions of the sources, the sensors, and how does the sound propagate between them. The most commonly used *free field* geometrical model assumes that the microphone configuration is perfectly known and that the sound propagates as a single plane wave from each source to each sensor (no reflection or interference). This model is not valid in realistic scenarios where the environment may be unknown, cluttered, dynamic, and include multiple sources, diffuse sounds, noise and/or reverberations. Such difficulties critical hinders sound source separation and localization tasks. In some ongoing work, we showed that the knowledge of a few early acoustic echoes significantly improve sound source separation performance over the free-field model.

Recently, two directions for advanced audio geometry estimation have emerged and were investigated in our team. The first one is physics-driven [46]. This approach explicitly solves the wave propagation equation in a given simplified yet realistic environment assuming that only few sound sources are present, in order to recover the positions of sources, sensors, or even some of the wall absorption properties. Encouraging results were obtained in simulated settings, including "hearing behind walls" [76]. However, these methods rely on approximate models and on partial knowledge of the system (e.g. room dimensions), limiting their real-world applicability so far. The second direction is data-driven. It uses machine learning to bypass the use of a physical

model by directly estimating a mapping from acoustic features to source positions, using training data obtained in a real room [65], [67]. These methods can in principle work in arbitrarily complex environments, but they require carefully annotated training datasets. Since obtaining such data is time consuming, the methods are usually working well for one specific room and setup, and are hard to generalize in practice.

We proposed a new paradigm that aims at making the best of physics-driven and data-driven approaches, referred to as *virtually acoustic space travelling* (VAST) [22], [30]. The idea is to use a physics-based room-acoustic simulator to generate arbitrary large datasets of room-impulse responses corresponding to various acoustic environments, adapted to the physical audio system at hand. We demonstrated that mappings learned from these data could potentially be used to not only estimate the 3D position of a source but also some acoustical properties of the room [30]. We also showed that a virtually-learned mapping could robustly localize sound sources from real-world binaural input, which is the first result of this kind in audio source localization [22]. The starting PhD thesis of Diego Di Carlo aims at applying the VAST framework to the blind estimation of acoustic echoes. The ultimate goal is to use these estimates to recover partial acoustic properties of the scene and enhance audio signal processing methods.

## 7.5. Music Content Processing and Information Retrieval

Music structure, music language modeling, System & Contrast model, complexity

Current work developed in our research group in the domain of music content processing and information retrieval explore various information-theoretic frameworks for music structure analysis and description [51], in particular the System & Contrast model [1].

#### 7.5.1. Tensor-based Representation of Sectional Units in Music

Participants: Corentin Guichaoua, Frédéric Bimbot.

Following Kolmogorov's complexity paradigm, modeling the structure of a musical segment can be addressed by searching for the compression program that describes as economically as possible the musical content of that segment, within a given family of compression schemes.

In this general framework, packing the musical data in a tensor-derived representation enables to decompose the structure into two components : (i) the shape of the tensor which characterizes the way in which the musical elements are arranged in an n-dimensional space and (ii) the values within the tensor which reflect the content of the musical segment and minimize the complexity of the relations between its elements.

This approach has been studied in the context of Corentin Guichaoua's PhD [11] where a novel method for the inference of musical structure based on the optimisation of a tensorial compression criterion has been designed and experimented.

This tensorial compression criterion exploits the redundancy resulting from repetitions, similarities, progressions and analogies within musical segments in order to pack musical information observed at different timescales in a single n-dimensional object.

The proposed method has been introduced from a formal point of view and has been related to the System & Constrast Model [1] as a extension of that model to hypercubic tensorial patterns and their deformations.

From the experimental point of view, the method has been tested on 100 pop music pieces (RWC Pop database) represented as chord sequences, with the goal to locate the boundaries of structural segments on the basis of chord grouping by minimizing the complexity criterion. The results have clearly established the relevance of the tensorial compression approach, with F-measure scores reaching 70 %

#### 7.5.2. Modeling music by polytopic graphs of latent relations

Participants: Corentin Louboutin, Frédéric Bimbot.

The musical content observed at a given instant within a music segment obviously tends to share privileged relationships with its immediate past, hence the sequential perception of the music flow. But local music content also relates with distant events which have occurred in the longer term past, especially at instants which

are metrically homologous (in previous bars, motifs, phrases, etc.) This is particularly evident in strongly "patterned" music, such as pop music, where recurrence and regularity play a central role in the design of cyclic musical repetitions, anticipations and surprises.

The web of musical elements can be described as a Polytopic Graph of Latent Relations (PGLR) which models relationships developing predominantly between homologous elements within the metrical grid.

For regular segments the PGLR lives on an n-dimensional cube(square, cube, tesseract, etc...), n being the number of scales considered simultaneously in the multiscale model. By extension, the PGLR can be generalized to a more or less regular n-dimensional polytopes.

Each vertex in the polytope corresponds to a low-scale musical element, each edge represents a relationship between two vertices and each face forms an elementary system of relationships.

The estimation of the PGLR structure of a musical segment can be obtained computationally as the joint estimation of the description of the polytope, the nesting configuration of the graph over the polytope (reflecting the flow of dependencies and interactions between the elements within the musical segment) and the set of relations between the nodes of the graph, with potentially multiple possibilities.

If musical elements are chords, relations can be inferred by minimal transport [85] defined as the shortest displacement of notes, in semitones, between a pair of chords. Other chord representations and relations are possible, as studied in [33] where the PGLR approach is presented conceptually and algorithmically, together with an extensive evaluation on a large set of chord sequences from the RWC Pop corpus (100 pop songs).

Specific graph configurations, called Primer Preserving Permutations (PPP) are extensively studied in [32] and are related to 6 main redundant sequences which can be viewed as canonical multiscale structural patterns.

These results illustrate the efficiency of the proposed model in capturing structural information within musical data and is currently being explored on melodic sequences and rythmic patterns.

#### 7.5.3. Regularity Constraints for the Fusion of Music Structure Segmentation System Participant: Frédéric Bimbot.

#### Main collaborations Gabriel Sargent (LinkMedia Inria project-team, Rennes)

Music structure estimation has become a central topic within the field of Music Information Retrieval. Indeed, as music is a highly structured information stream, knowledge of how a music piece is organized represents a key challenge to enhance the management and exploitation of large music collections.

Former work carried out in our group [94] has illustrated the benefits that can be expected from a regularity constraint on the structural segmentation of popular music pieces : a constraint which favors structural segments of comparable size provides a better conditioning of the boundary estimation process.

As a further investigation, we have explored the benefits of the regularity constraint as an efficient way for combining the outputs of a selection of systems presented at MIREX between 2010 and 2015. These experiments have yielded a level of performance which is competitive to that of the state-of-the-art on the "MIREX10" dataset (100 J-Pop songs from the RWC database) [21].

# 8. Bilateral Contracts and Grants with Industry

# 8.1. Bilateral Grants with Industry

8.1.1. CIFRE contract with Technicolor R&I France on Very large scale visual comparison Participants: Rémi Gribonval, Himalaya Jain.

> Duration: 3 years (2015-2018) Research axis: 3.1.2 Partners: Technicolor R&I France; Inria-Rennes Funding: Technicolor R&I France; ANRT

The grand goal of this thesis is to design, analyze and test new tools to allow large-scale comparison of high-dimensional visual signatures. Leveraging state of the art visual descriptors, the objective is to obtain new compact codes for visual representations, exploiting sparsity and learning, so that they can be stored and compared in an efficient, yet meaningful, way.

# 9. Partnerships and Cooperations

# 9.1. National Initiatives

### 9.1.1. Labex Comin Labs projects

CominLabs is a Laboratoire d'Excellence funded by the PIA (Programme Investissements d'Avenir) in the broad area of telecommunications.

#### 9.1.1.1. HEMISFER

#### Participant: Rémi Gribonval.

Acronym: HYBRID (Hybrid Eeg-MrI and Simultaneous neuro-feedback for brain Rehabilitation)

http://www.hemisfer.cominlabs.ueb.eu/

Research axis: 3.1

CominLabs partners : VISAGES, HYBRID and PANAMA Inria project-teams;

External partners : EA 4712 team from University of Rennes I; ATHENA Inria project-team, Sophia-Antipolis;

Coordinator: Christian Barillot, VISAGES Inria project-team

Description: The goal of HEMISFER is to make full use of neurofeedback paradigm in the context of rehabilitation and psychiatric disorders. The major breakthrough will come from the use of a coupling model associating functional and metabolic information from Magnetic Resonance Imaging (fMRI) to Electro-encephalography (EEG) to "enhance" the neurofeedback protocol. We propose to combine advanced instrumental devices (Hybrid EEG and MRI platforms), with new man-machine interface paradigms (Brain computer interface and serious gaming) and new computational models (source separation, sparse representations and machine learning) to provide novel therapeutic and neuro-rehabilitation paradigms in some of the major neurological and psychiatric disorders of the developmental and the aging brain (stroke, attention-deficit disorder, language disorders, treatment-resistant mood disorders, ...).

Contribution of PANAMA: PANAMA, in close cooperation with the VISAGES team, contributes to a coupling model between EEG and fMRI considered as a joint inverse problem addressed with sparse regularization. By combining both modalities, one expects to achieve a good reconstruction both in time and space. This new imaging technique will then be used for improving neurofeedback paradigms in the context of rehabilitation and psychiatric disorders, which is the final purpose of the HEMISFER project.

#### 9.1.1.2. TEPN

Participant: Rémi Gribonval.

Acronym: TEPN (Toward Energy Proportional Networks)

http://www.tepn.cominlabs.ueb.eu/

Research axis: 3.1

CominLabs partners : IRISA OCIF - Telecom Bretagne; IETR SCN; IETR SCEE; PANAMA Inria project-team

Coordinator: Nicolas Montavont, IRISA OCIF - Telecom Bretagne

Description: As in almost all areas of engineering in the past several decades, the design of computer and network systems has been aimed at delivering maximal performance without regarding to the energy efficiency or the percentage of resource utilization. The only places where this tendency was questioned were battery-operated devices (such as laptops and smartphones) for which the users accept limited (but reasonable) performance in exchange for longer use periods. Even though the end users make such decisions on a daily basis by checking their own devices, they have no way of minimizing their energy footprint (or conversely, optimize the network resource usage) in the supporting infrastructure. Thus, the current way of dimensioning and operating the infrastructure supporting the user services, such as cellular networks and data centers, is to dimension for peak usage. The problem with this approach is that usage is rarely at its peak. The overprovisioned systems are also aimed at delivering maximal performance, with energy efficiency being considered as something desired, but non-essential. This project aims at making the network energy consumption proportional to the actual charge of this network (in terms of number of served users, or requested bandwidth). An energy proportional network can be designed by taking intelligent decisions (based on various constraints and metrics) into the network such as switching on and off network components in order to adapt the energy consumption to the user needs. This concept can be summarized under the general term of Green Cognitive Network Approach.

Contribution of PANAMA: PANAMA, in close cooperation with the SCEE team at IETR (thesis of Marwa Chafii, 2016), focuses on the design of new waveforms for multi carrier systems with reduced Peak to Average Power Ratio (PAPR).

#### 9.1.2. ANR INVATE project with IRT b<>com, Rennes

Participants: Rémi Gribonval, Nancy Bertin, Mohammed Hafsati.

Thesis on 3D audio scene decomposition for interactive navigation

Duration: 3 years (2016-2019)

Research axis: 3.2.2

Partners: IRT b<>com; Inria-Rennes; IRISA

Funding: ANR INVATE project (PIA)

The objective of this thesis is to develop tools to analyze audio scenes in order to identify, locate, and extract the sources present in the scene to re-spatialize them according to the user head orientation and the movement of the user in the targeted virtual scene.

#### 9.1.3. ANR OATMIL project

Participants: Rémi Gribonval, Antoine Chatalic.

Duration: 4 years (2017-2021)

Acronym: OATMIL (Bringing Optimal Transport and Machine Learning Together) http://people.irisa.fr/Nicolas.Courty/OATMIL/

Research Axis 3.1

Partners: Obelix team and PANAMA Inria project-team, IRISA; LITIS, Rouen; Lagrange Laboratory, Nice; Technicolor R&I France, Rennes.

Coordinator: Nicolas Courty (Obelix team)

Description: The OATMIL project will propose novel concepts, methodologies, and new tools for exploiting large data collections. This will result from a cross-fertilization of fundamental tools and ideas from optimal transport (OT) and machine learning (ML). The main objective of OATMIL is to develop new techniques for large-scale machine learning, encompassing adaptability, scalability, and robustness, by a cross-fertilization of ideas coming from OT and ML. This cross-fertilization leads to two complementary scientific challenges : bringing OT to ML and bringing ML to OT.

Contribution of PANAMA: PANAMA will explore the use of dimension-reduction with sketching strategies in the context compressive optimal transport. Funding: ANR

#### 9.1.4. OSEO-FUI: voiceHome

Participants: Nancy Bertin, Frédéric Bimbot, Romain Lebarbenchon, Ewen Camberlein.

Duration: 3 years (2015-2017)

Research axis: 3.2

Partners: voicebox (formerly known as onMobile), Delta Dore, eSoftThings, Orange, Technicolor R&I France, LOUSTIC, Inria Nancy

Coordinator: voicebox

Description: The goal of the project is to design and implement a multi-channel voice interface for smart home and multimedia (set-top-box) appliances.

Contributions of PANAMA are focused on audio source localization and separation with distant microphones in real environments. In both cases, the issue of energy frugality is central and strongly constrains the available resources. This cooperation, which reached its end in November 2017, allowed us to make progress towards operational low-resource audio source localization and separation schemes, to disseminate software, collected data and scientific results, and to identify new research and development perspectives in adaptive microphone array processing for fast and robust audio scene analysis.

# 9.2. International Initiatives

## 9.2.1. Inria International Partners

#### 9.2.1.1. Informal International Partners

PANAMA has strong recurrent collaborations with the LTS2 lab at EPFL, the Center for Digital Music at Queen Mary University of London, the Institute for Digital Communications at the University of Edimburgh, and the Institute for Mathematics of the Postdam University.

## 9.3. International Research Visitors

#### 9.3.1. Visits of International Scientists

- Flavio Castro Alves Teixeira, in May-June 2017, Post-doc, University of Innsbruck, Austria
- Pierre Vandergheynst, in June-July 2017, Professor of Signal and Image Processing, EPFL (Chaire Internationale Inria), Lausanne, Switzerland
- Gilles Blanchard, in September 2017, Professor, University of Potsdam, Germany
- Mike Davies, in October 2017, Professor, University of Edimburgh, UK
- Jérémy Cohen, in November 2017, Post-doc, University of Mons, Belgium
- Andreas Loukas, in December 2017, Post-doc, EPFL, Lausanne, Switzerland

9.3.1.1. Internships

- Helena Peic Tukuljac, from October to December 2017, PhD Student at EPFL, Lausanne, Switzerland
- Martin Strauss, from October to December 2017, M1 student, Friedrich-Alexander University, Erlangen, Germany

# **10.** Dissemination

# **10.1. Promoting Scientific Activities**

Antoine Deleforge was elected member of the IEEEAudio and Acoustic Signal Processing Technical Committee. Rémi Gribonval is a member of the IEEE Technical Committee on Signal Processing Theory and Methods (2012–2017), and a member of the Awards sub-committee.

Rémi Gribonval is a member of the program committee of the GRETSI.

Rémi Gribonval is a member of the EURASIP Special Area Team (SAT) on Signal and Data Analytics for Machine Learning (SiG-DML) since 2015.

Rémi Gribonval has been a member of the Steering Committee of the SPARS international workshop (chairman until 2013) until 2017.

Frédéric Bimbot is the Head of the "Digital Signals and Images, Robotics" department in IRISA (UMR 6074).

Frédéric Bimbot is a member of the International Advisory Council of ISCA (International Speech Communication Association).

Rémi Gribonval and Frédéric Bimbot are the scientific coordinators of the Science and Music Day (Journée Science et Musique) organized by IRISA.

Nancy Bertin and Frédéric Bimbot are the coordinators of the Science and Music Young Researcher Award (Prix Jeune Chercheur Science et Musique).

Antoine Deleforge organized and will co-chair with Ivan Dokmanic and Robin Schleibler a special session on "Geometry-Aware Auditory Scene Analysis" at ICASSP 2018, Calgary, Canada.

N. Bertin, A. Deleforge, and R. Gribonval co-organized a special session entitled "From Source Position to Room Properties : Learning Methods for Audio Scene Geometry Estimation", at the 2017 LVA/ICA Conference, Grenoble, February 2017

Antoine Deleforge is Area Chair in Bayesian Inference for the 14th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2018).

R. Gribonval was an invited speaker at the Learning Theory Workshop during the conference on Foundations of Computational Mathematics (Barcelona, July 2017) and the Les Houches Workshop on Statistical Physics (Les Houches, February 2017).

R. Gribonval has been a member of the jury of the GDR ISIS / GRETSI / Club EAA thesis prize in signal and image processing from 2014 to 2017.

R. Gribonval is the scientific organizer of the 13th Peyresq summer school in signal and image processing, jointly organized in July 2018 by GRETSI and GDR ISIS.

R. Gribonval has joined in 2017 the Scientific Advisory Board of the Acoustics Research Institute from the Austrian Academy of Sciences in Vienna.

## **10.2.** Teaching - Supervision - Juries

#### 10.2.1. Teaching

Bachelor : N. Bertin, "Discovery of selected topics in audio signal processing research", 6 hours, L3, École Supérieure de Réalisation Audiovisuelle (ESRA), France.

Master : N. Bertin, "Vocal and Audio Interactions", 4 hours, M2, Université Rennes 1, France.

Master : N. Bertin, "Fundamentals of Signal Processing", 24 hours, M1, Ecole Normale Supérieure (ENS) de Bretagne, Rennes, France.

Master : N. Bertin, "Sparse representations and compressive sensing", 15 hours, M2, Ecole Nationale de la Statistique et de l'Analyse de l'Information (ENSAI), Rennes, France.

Master : N. Bertin, "Sparsity in Signal and Image Processing", 12 hours, M2, Institut National des Sciences Appliquées (INSA) de Rennes, France.

Master : R. Gribonval, "High dimensional statistical learning", 12 hours, M2, Université Rennes 1, France.

Master: R. Gribonval, coordination of the HDL module "High dimensional statistical learning" within the SIF M2, 20 hours, Université Rennes 1, France.

Master : R. Gribonval, "Sparsity in Signal and Image Processing", 4 hours, M2, Institut National des Sciences Appliquées (INSA) de Rennes, France.

Master : R. Gribonval, coordination of the module "Sparsity in Signal and Image Processing", 48 hours, M2, Institut National des Sciences Appliquées (INSA) de Rennes, France.

Bachelor : A. Deleforge, "Discovery of selected topics in audio signal processing research", 6 hours, L3, École Supérieure de Réalisation Audiovisuelle (ESRA), France.

Master : A. Deleforge, "Sparsity in Signal and Image Processing", 4 hours, M2, Institut National des Sciences Appliquées (INSA) de Rennes, France.

Master : A. Deleforge, "Vocal and Audio Interaction", 6 hours, M2, Université Rennes 1, France.

Master : F. Bimbot, coordination of the VAI module "Vocal and Acoustic Interactions" within the SIF M2, 20 hours, Université de Rennes 1, France

# **10.3.** Popularization

#### 10.3.1. Journée Science et Musique

**Participants:** Antoine Deleforge, Rémi Gribonval, Frédéric Bimbot, Romain Lebarbenchon, Clément Gaultier, Nancy Bertin, Ewen Camberlein, Stéphanie Lemaile, Corentin Louboutin, Corentin Guichaoua, Cássio Fraga Dantas, Valentin Gillot, Antoine Chatalic, Yann Traonmilin.

with contributions and support from: Valérie Gouranton, Ronan Gaugne, Evelyne Orain, Agnès Cottais, Catherine Jacques-Orban and many more.

PANAMA coordinated the organization of a public event called "Journée Science et Musique" ("Music and Science Day"). This yearly event organized by the METISS/ PANAMA Team since 2011 aims at sharing with the wide audience the latest innovations and research projects in music. The motivation for hosting this event is to explain and promote the technology behind audio-processing that people face in their daily lives. The event is free to everyone and people have the possibility to attend talks by selected speakers or meet numerous experts that demonstrate current projects in which people can interactively participate. Edition 2017 hosted more than 800 visitors and was a partner of the "Festival des Sciences" and "Festival Maintenant" in Rennes.

# **11. Bibliography**

## Major publications by the team in recent years

- [1] F. BIMBOT, E. DERUTY, G. SARGENT, E. VINCENT. System & Contrast : A Polymorphous Model of the Inner Organization of Structural Segments within Music Pieces, in "Music Perception", June 2016, vol. 33, n<sup>o</sup> 5, p. 631-661 [DOI: 10.1525/MP.2016.33.5.631], https://hal.inria.fr/hal-01188244.
- [2] N. DUONG, E. VINCENT, R. GRIBONVAL.Under-determined reverberant audio source separation using a fullrank spatial covariance model, in "IEEE Transactions on Audio, Speech and Language Processing", July 2010, vol. 18, n<sup>o</sup> 7, p. 1830–1840 [DOI: 10.1109/TASL.2010.2050716], https://hal.inria.fr/inria-00541865.
- [3] R. GRIBONVAL, K. SCHNASS.Dictionary Identification Sparse Matrix-Factorisation via ℓ\_1-Minimisation, in "IEEE Transactions on Information Theory", July 2010, vol. 56, n<sup>o</sup> 7, p. 3523–3539 [DOI: 10.1109/TIT.2010.2048466], https://hal.archives-ouvertes.fr/hal-00541297.

- [4] D. K. HAMMOND, P. VANDERGHEYNST, R. GRIBONVAL. Wavelets on graphs via spectral graph theory, in "Applied and Computational Harmonic Analysis", March 2011, vol. 30, n<sup>o</sup> 2, p. 129–150 [DOI: 10.1016/J.ACHA.2010.04.005], https://hal.inria.fr/inria-00541855.
- [5] S. KITIĆ, L. ALBERA, N. BERTIN, R. GRIBONVAL. Physics-driven inverse problems made tractable with cosparse regularization, in "IEEE Transactions on Signal Processing", January 2016, vol. 64, n<sup>o</sup> 2, p. 335-348 [DOI: 10.1109/TSP.2015.2480045], https://hal.inria.fr/hal-01133087.
- [6] S. KITIĆ. *Cosparse regularization of physics-driven inverse problems*, IRISA, Inria Rennes, November 2015, https://hal.archives-ouvertes.fr/tel-01237323.
- [7] C. LOUBOUTIN, F. BIMBOT. Modeling the multiscale structure of chord sequences using polytopic graphs, in "18th International Society for Music Information Retrieval Conference", Suzhou, China, October 2017, https://hal.archives-ouvertes.fr/hal-01653455.
- [8] S. NAM, M. E. DAVIES, M. ELAD, R. GRIBONVAL. *The Cosparse Analysis Model and Algorithms*, in "Applied and Computational Harmonic Analysis", 2013, vol. 34, n<sup>o</sup> 1, p. 30–56, Preprint available on arXiv since 24 Jun 2011 [*DOI* : 10.1016/J.ACHA.2012.03.006], http://hal.inria.fr/inria-00602205.
- [9] A. OZEROV, E. VINCENT, F. BIMBOT. A General Flexible Framework for the Handling of Prior Information in Audio Source Separation, in "IEEE Transactions on Audio, Speech and Language Processing", May 2012, vol. 20, n<sup>0</sup> 4, p. 1118 - 1133, 16, http://hal.inria.fr/hal-00626962.
- [10] E. VINCENT, N. BERTIN, R. GRIBONVAL, F. BIMBOT. From blind to guided audio source separation, in "IEEE Signal Processing Magazine", December 2013, http://hal.inria.fr/hal-00922378.

## **Publications of the year**

#### **Doctoral Dissertations and Habilitation Theses**

- [11] C. GUICHAOUA. Compression models and complexity criteria for the description and the inference of music structure, Université Rennes 1, September 2017, https://tel.archives-ouvertes.fr/tel-01687054.
- [12] N. KERIVEN. Sketching for large-scale learning of mixture models, Université Rennes 1, October 2017, https:// tel.archives-ouvertes.fr/tel-01620815.

#### **Articles in International Peer-Reviewed Journal**

- [13] H. BECKER, L. ALBERA, P. COMON, J.-C. NUNES, R. GRIBONVAL, J. FLEUREAU, P. GUILLOTEL, I. MERLET.SISSY: an efficient and automatic algorithm for the analysis of EEG sources based on structured sparsity, in "NeuroImage", August 2017, vol. 157, p. 157-172 [DOI: 10.1016/J.NEUROIMAGE.2017.05.046], http://www.hal.inserm.fr/inserm-01617155.
- [14] M. CHAFII, J. PALICOT, R. GRIBONVAL. Wavelet modulation: An alternative modulation with low energy consumption, in "Comptes Rendus Physique", 2017, vol. 18, n<sup>o</sup> 2, p. 156-167 [DOI: 10.1016/J.CRHY.2016.11.010], https://hal.archives-ouvertes.fr/hal-01445465.
- [15] C. F. DANTAS, M. N. DA COSTA, R. DA ROCHA LOPES.Learning Dictionaries as a sum of Kronecker products, in "IEEE Signal Processing Letters", May 2017, vol. 24, n<sup>o</sup> 5, p. 559 - 563 [DOI: 10.1109/LSP.2017.2681159], https://hal.inria.fr/hal-01672349.

- [16] V. DROUARD, R. HORAUD, A. DELEFORGE, S. BA, G. EVANGELIDIS. *Robust Head-Pose Estimation Based on Partially-Latent Mixture of Linear Regressions*, in "IEEE Transactions on Image Processing", March 2017, vol. 26, n<sup>o</sup> 3, p. 1428 1440, https://arxiv.org/abs/1603.09732 [DOI: 10.1109/TIP.2017.2654165], https:// hal.inria.fr/hal-01413406.
- [17] L. LE MAGOAROU, R. GRIBONVAL, N. TREMBLAY. Approximate fast graph Fourier transforms via multilayer sparse approximations, in "IEEE transactions on Signal and Information Processing over Networks", June 2017, 15, https://arxiv.org/abs/1612.04542 [DOI: 10.1109/TSIPN.2017.2710619], https://hal.inria. fr/hal-01416110.
- [18] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE. *Aural servo: sensor-based control from robot audition*, in "IEEE Transactions on Robotics", 2018, p. 1-14, https://hal.inria.fr/hal-01694366.
- [19] E. PERTHAME, F. FORBES, A. DELEFORGE. Inverse regression approach to robust nonlinear high-tolow dimensional mapping, in "Journal of Multivariate Analysis", January 2018, vol. 163, p. 1 - 14 [DOI: 10.1016/J.JMVA.2017.09.009], https://hal.inria.fr/hal-01652011.
- [20] G. PUY, M. E. DAVIES, R. GRIBONVAL. *Recipes for stable linear embeddings from Hilbert spaces to*  $\mathbb{R}^m$ , in "IEEE Transactions on Information Theory", 2017, https://arxiv.org/abs/1509.06947 Submitted in 2015 [*DOI*: 10.1109/TIT.2017.2664858], https://hal.inria.fr/hal-01203614.
- [21] G. SARGENT, F. BIMBOT, E. VINCENT. Estimating the structural segmentation of popular music pieces under regularity constraints, in "IEEE/ACM Transactions on Audio, Speech, and Language Processing", 2017, https://hal.inria.fr/hal-01403210.

#### **Invited Conferences**

- [22] C. GAULTIER, S. KATARIA, A. DELEFORGE.VAST : The Virtual Acoustic Space Traveler Dataset, in "International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)", Grenoble, France, February 2017, https://arxiv.org/abs/1612.06287, https://hal.archives-ouvertes.fr/hal-01416508.
- [23] N. KERIVEN, N. TREMBLAY, Y. TRAONMILIN, R. GRIBONVAL. Compressive K-means, in "IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)", New Orleans, United States, March 2017, https://arxiv.org/abs/1610.08738, https://hal.inria.fr/hal-01386077.

#### **International Conferences with Proceedings**

- [24] F. ARGELAGUET, M. DUCOFFE, A. LÉCUYER, R. GRIBONVAL. Spatial and rotation invariant 3D gesture recognition based on sparse representation, in "IEEE Symposium on 3D User Interfaces", Los Angeles, United States, March 2017, p. 158 - 167 [DOI: 10.1109/3DUI.2017.7893333], https://hal.inria.fr/hal-01625128.
- [25] E. BYRNE, R. GRIBONVAL, P. SCHNITER. Sketched Clustering via Hybrid Approximate Message Passing, in "Asilomar Conference on Signals, Systems, and Computers", Pacific Grove, California, United States, October 2017, https://hal.inria.fr/hal-01650160.
- [26] A. DELEFORGE, Y. TRAONMILIN. Phase Unmixing : Multichannel Source Separation with Magnitude Constraints, in "2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)", New Orleans, United States, March 2017, https://arxiv.org/abs/1609.09744, https://hal.inria.fr/hal-01372418.

- [27] A. DRÉMEAU, A. DELEFORGE. Phase Retrieval with a Multivariate Von Mises Prior: From a Bayesian Formulation to a Lifting Solution, in "ICASSP 2017 - 42nd IEEE International Conference on Acoustics, Speech and Signal Processing", New Orleans, United States, March 2017, p. 1-5 [DOI: 10.1109/ICASSP.2017.7953027], https://hal.archives-ouvertes.fr/hal-01653732.
- [28] C. GAULTIER, S. KITIĆ, N. BERTIN, R. GRIBONVAL.AUDASCITY: AUdio Denoising by Adaptive Social CosparsITY, in "25th European Signal Processing Conference (EUSIPCO)", Kos, Greece, August 2017, https://hal.inria.fr/hal-01540945.
- [29] H. JAIN, J. ZEPEDA, P. PÉREZ, R. GRIBONVAL.SUBIC: A supervised, structured binary code for image search, in "The IEEE International Conference on Computer Vision (ICCV)", Venise, Italy, IEEE, October 2017, https://arxiv.org/abs/1708.02932 - Accepted at ICCV 2017 (Spotlight) [DOI: 10.1109/ICCV.2017.96], https://hal.inria.fr/hal-01683390.
- [30] S. KATARIA, C. GAULTIER, A. DELEFORGE.*Hearing in a shoe-box : binaural source position and wall absorption estimation using virtually supervised learning*, in "2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)", New-Orleans, United States, March 2017, https://arxiv.org/abs/1609.09747, https://arxiv.org/abs/1609.09747
- [31] L. LE MAGOAROU, N. TREMBLAY, R. GRIBONVAL. Analyzing the Approximation Error of the Fast Graph Fourier Transform, in "ASILOMAR conference on Signals, Systems, and Computers", Monterey, California, United States, October 2017, https://hal.archives-ouvertes.fr/hal-01627434.
- [32] C. LOUBOUTIN, F. BIMBOT. Modeling the multiscale structure of chord sequences using polytopic graphs, in "18th International Society for Music Information Retrieval Conference", Suzhou, China, October 2017, https://hal.archives-ouvertes.fr/hal-01653455.
- [33] C. LOUBOUTIN, F. BIMBOT. Polytopic Graph of Latent Relations: A Multiscale Structure Model for Music Segments, in "6th International Conference on Mathematics and Computation in Music (MCM 2017)", Mexico City, Mexico, O. A. AGUSTÍN-AQUINO, E. LLUIS-PUEBLA, M. MONTIEL (editors), Lecture Notes in Computer Science book series, Springer, June 2017, vol. 10527, https://hal.archives-ouvertes.fr/hal-01653445.
- [34] S. NOORZADEH, P. MAUREL, T. OBERLIN, R. GRIBONVAL, C. BARILLOT. Multi-modal EEG and fMRI Source Estimation Using Sparse Constraints, in "MICCAI 2017 - 20th International Conference on Medical Image Computing and Computer Assisted Intervention", Quebec, Canada, September 2017 [DOI: 10.1007/978-3-319-66182-7\_51], https://hal.inria.fr/hal-01586495.

#### **National Conferences with Proceeding**

[35] C. FRAGA DANTAS, R. GRIBONVAL. *Dynamic Screening with Approximate Dictionaries*, in "XXVIème colloque GRETSI", Juan-les-Pins, France, September 2017, https://hal.inria.fr/hal-01598021.

#### **Scientific Popularization**

[36] K. DÉGUERNEL, N. LIBERMANN, E. VINCENT. La musique comme une langue, March 2017, Commission française pour l'enseignement des mathématiques livret "Mathématiques et langages - Panorama du thème", https://hal.inria.fr/hal-01485209.

#### **Other Publications**

- [37] I. DOKMANIĆ, R. GRIBONVAL.Beyond Moore-Penrose Part I: Generalized Inverses that Minimize Matrix Norms, July 2017, https://arxiv.org/abs/1706.08349 - working paper or preprint, https://hal.inria.fr/hal-01547183.
- [38] I. DOKMANIĆ, R. GRIBONVAL. Beyond Moore-Penrose Part II: The Sparse Pseudoinverse, July 2017, https:// arxiv.org/abs/1706.08701 - working paper or preprint, https://hal.inria.fr/hal-01547283.
- [39] C. FRAGA DANTAS, R. GRIBONVAL, R. R. LOPES, M. N. DA COSTA.Learning Dictionaries as Sums of Kronecker Products, June 2017, SPARS 2017 - Signal Processing with Adaptive Sparse Structured Representations workshop, https://hal.inria.fr/hal-01514044.
- [40] C. GAULTIER, N. BERTIN, S. KITIĆ, R. GRIBONVAL. A modeling and algorithmic framework for (non)social (co)sparse audio restoration, November 2017, https://arxiv.org/abs/1711.11259 - working paper or preprint, https://hal.inria.fr/hal-01649261.
- [41] C. GAULTIER, S. KITIĆ, N. BERTIN, R. GRIBONVAL. Cosparse Denoising: The Importance of Being Social, June 2017, The Signal Processing with Adaptive Sparse Structured Representations (SPARS) workshop, https://hal.inria.fr/hal-01510710.
- [42] R. GRIBONVAL, G. BLANCHARD, N. KERIVEN, Y. TRAONMILIN. Compressive Statistical Learning with Random Feature Moments, December 2017, https://arxiv.org/abs/1706.07180 - Main novelties compared to version 1: improved concentration bounds, improved sketch sizes for compressive k-means and compressive GMM that now scale linearly with the ambient dimension, https://hal.inria.fr/hal-01544609.
- [43] R. GRIBONVAL, H. JAIN, J. ZEPEDA, P. PÉREZ.Learning a Complete Image Indexing Pipeline, January 2018, https://arxiv.org/abs/1712.04480 - working paper or preprint, https://hal.inria.fr/hal-01683385.
- [44] N. KERIVEN, A. BOURRIER, R. GRIBONVAL, P. PÉREZ. Sketching for Large-Scale Learning of Mixture Models, May 2017, https://arxiv.org/abs/1606.02838 - working paper or preprint, https://hal.inria.fr/hal-01329195.
- [45] N. KERIVEN, A. DELEFORGE, A. LIUTKUS.Blind Source Separation Using Mixtures of Alpha-Stable Distributions, November 2017, https://arxiv.org/abs/1711.04460 - working paper or preprint, https://hal.inria. fr/hal-01633215.
- [46] S. KITIĆ, S. BENSAID, L. ALBERA, N. BERTIN, R. GRIBONVAL. Versatile and scalable cosparse methods for physics-driven inverse problems, March 2017, working paper or preprint, https://hal.inria.fr/hal-01496767.
- [47] Y. TRAONMILIN, N. KERIVEN, R. GRIBONVAL, G. BLANCHARD. Spikes super-resolution with random Fourier sampling, June 2017, SPARS workshop 2017, https://hal.archives-ouvertes.fr/hal-01509863.
- [48] Y. TRAONMILIN, G. PUY, R. GRIBONVAL, M. E. DAVIES. *Compressed sensing in Hilbert spaces*, July 2017, https://arxiv.org/abs/1702.04917 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01469134.

#### **References in notes**

[49] A. ADLER, V. EMIYA, M. G. JAFARI, M. ELAD, R. GRIBONVAL, M. D. PLUMBLEY. Audio Inpainting, in "IEEE Transactions on Audio, Speech and Language Processing", March 2012, vol. 20, n<sup>o</sup> 3, p. 922 - 932 [DOI: 10.1109/TASL.2011.2168211], http://hal.inria.fr/inria-00577079.

- [50] L. ALBERA, S. KITIĆ, N. BERTIN, G. PUY, R. GRIBONVAL.Brain source localization using a physicsdriven structured cosparse representation of EEG signals, in "2014 IEEE International Workshop on Machine Learning for Signal Processing", Reims, France, September 2014, 6, https://hal.archives-ouvertes.fr/hal-01027609.
- [51] F. BIMBOT. Towards an Information-Theoretic Framework for Music Structure, 2016, p. 167-168, Invited talk at Dagstuhl Seminar 16092 on Computational Music Structure Analysis (Ed. : M. Müller, E. Chew, J.P. Bello) [DOI: 10.4230/DAGREP.6.2.147], https://hal.archives-ouvertes.fr/hal-01421013.
- [52] A. BONNEFOY, V. EMIYA, L. RALAIVOLA, R. GRIBONVAL.Dynamic Screening: Accelerating First-Order Algorithms for the Lasso and Group-Lasso, in "IEEE Transactions on Signal Processing", 2015, vol. 63, n<sup>o</sup> 19, 20 [DOI: 10.1109/TSP.2015.2447503], https://hal.archives-ouvertes.fr/hal-01084986.
- [53] A. BOURRIER. Compressed sensing and dimensionality reduction for unsupervised learning, Université Rennes 1, May 2014, https://tel.archives-ouvertes.fr/tel-01023030.
- [54] A. BOURRIER, R. GRIBONVAL, P. PÉREZ. Compressive Gaussian Mixture Estimation, in "Signal Processing with Adaptive Sparse Structured Representations (SPARS) 2013", Switzerland, July 2013, http://hal.inria.fr/ hal-00811819.
- [55] A. BOURRIER, R. GRIBONVAL, P. PÉREZ. Compressive Gaussian Mixture Estimation, in "ICASSP 38th International Conference on Acoustics, Speech, and Signal Processing", Vancouver, Canada, 2013, p. 6024-6028, http://hal.inria.fr/hal-00799896.
- [56] A. BOURRIER, R. GRIBONVAL, P. PÉREZ. *Estimation de mélange de Gaussiennes sur données compressées*, in "24ème Colloque Gretsi", France, September 2013, 191, http://hal.inria.fr/hal-00839579.
- [57] M. CHAFII.Study of a new multicarrier waveform with low PAPR, CentraleSupélec, October 2016, https://hal. archives-ouvertes.fr/tel-01399509.
- [58] M. CHAFII, J. PALICOT, R. GRIBONVAL, F. BADER. A Necessary Condition for Waveforms with Better PAPR than OFDM, in "IEEE Transactions on Communications", 2016 [DOI: 10.1109/TCOMM.2016.2584068], https://hal.inria.fr/hal-01128714.
- [59] M. CHAFII, J. PALICOT, R. GRIBONVAL. A PAPR upper bound of generalized waveforms for multi-carrier modulation systems, in "6th International Symposium on Communications, Control, and Signal Processing -ISCCSP 2014", Athènes, Greece, May 2014, p. 461 - 464 [DOI: 10.1109/ISCCSP.2014.6877913], https:// hal-supelec.archives-ouvertes.fr/hal-01072519.
- [60] M. CHAFII, J. PALICOT, R. GRIBONVAL. Closed-form Approximations of the PAPR Distribution for Multi-Carrier Modulation Systems, in "EUSIPCO 2014 - European Signal Processing Conference", Lisbonne, Portugal, September 2014, https://hal.inria.fr/hal-01054126.
- [61] M. CHAFII, J. PALICOT, R. GRIBONVAL.Closed-form approximations of the peak-to-average power ratio distribution for multi-carrier modulation and their applications, in "EURASIP Journal on Advances in Signal Processing", 2014, vol. 2014, n<sup>o</sup> 1, 121 [DOI : 10.1186/1687-6180-2014-121], https://hal.inria.fr/hal-01056153.

- [62] M. CHAFII, J. PALICOT, R. GRIBONVAL. L'optimalité de l'OFDM en termes de performance en PAPR, in "25ème Colloque Gretsi 2015", Lyon, France, September 2015, https://hal-supelec.archives-ouvertes.fr/hal-01165509.
- [63] M. CHAFII, J. PALICOT, R. GRIBONVAL. Dispositif de communication à modulation temps-fréquence adaptative, July 2016, n<sup>o</sup> Numéro de demande : 1656806 ; Numéro de soumission : 1000356937, https://hal.inria. fr/hal-01375661.
- [64] A. DELEFORGE, F. FORBES, S. BA, R. HORAUD.Hyper-Spectral Image Analysis With Partially Latent Regression and Spatial Markov Dependencies, in "IEEE Journal of Selected Topics in Signal Processing", 2015, vol. 9, n<sup>o</sup> 6, p. 1037–1048.
- [65] A. DELEFORGE, F. FORBES, R. HORAUD.Acoustic space learning for sound-source separation and localization on binaural manifolds, in "International journal of neural systems", 2015, vol. 25, n<sup>o</sup> 01, 1440003, https://hal.archives-ouvertes.fr/hal-00960796.
- [66] A. DELEFORGE, F. FORBES, R. HORAUD.*High-dimensional regression with gaussian mixtures and partiallylatent response variables*, in "Statistics and Computing", 2015, vol. 25, n<sup>o</sup> 5, p. 893–911.
- [67] A. DELEFORGE, R. HORAUD, Y. Y. SCHECHNER, L. GIRIN. Co-localization of audio sources in images using binaural features and locally-linear regression, in "IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)", 2015, vol. 23, n<sup>o</sup> 4, p. 718–731.
- [68] A. DELEFORGE, W. KELLERMANN. Phase-optimized K-SVD for signal extraction from underdetermined multichannel sparse mixtures, in "Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on", IEEE, 2015, p. 355–359.
- [69] M. A. GERZON. *Periphony: With-Height Sound Reproduction*, in "J. Audio Eng. Soc", 1973, vol. 21, n<sup>o</sup> 1, p. 2–10, http://www.aes.org/e-lib/browse.cfm?elib=2012.
- [70] R. GIRYES, S. NAM, M. ELAD, R. GRIBONVAL, M. E. DAVIES. Greedy-Like Algorithms for the Cosparse Analysis Model, in "Linear Algebra and its Applications", January 2014, vol. 441, p. 22–60, partially funded by the ERC, PLEASE project, ERC-2011-StG-277906 [DOI: 10.1016/J.LAA.2013.03.004], http://hal.inria. fr/hal-00716593.
- [71] N. KERIVEN, A. BOURRIER, R. GRIBONVAL, P. PÉREZ.Sketching for Large-Scale Learning of Mixture Models, in "2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016)", Shanghai, China, March 2016, https://hal.inria.fr/hal-01208027.
- [72] N. KERIVEN, R. GRIBONVAL. Compressive Gaussian Mixture Estimation by Orthogonal Matching Pursuit with Replacement, July 2015, SPARS 2015, https://hal.inria.fr/hal-01165984.
- [73] S. KITIĆ, N. BERTIN, R. GRIBONVAL. A review of cosparse signal recovery methods applied to sound source localization, in "Le XXIVe colloque Gretsi", Brest, France, September 2013, http://hal.inria.fr/hal-00838080.
- [74] S. KITIĆ, N. BERTIN, R. GRIBONVAL. Audio Declipping by Cosparse Hard Thresholding, August 2014, iTwist - 2nd international - Traveling Workshop on Interactions between Sparse models and Technology, https://hal.inria.fr/hal-00922497.

- [75] S. KITIĆ, N. BERTIN, R. GRIBONVAL.*Hearing behind walls: localizing sources in the room next door with cosparsity*, in "ICASSP IEEE International Conference on Acoustics, Speech, and Signal Processing", Florence, Italy, May 2014 [DOI: 10.1109/ICASSP.2014.6854168], https://hal.inria.fr/hal-00904779.
- [76] S. KITIĆ, N. BERTIN, R. GRIBONVAL. Hearing behind walls: localizing sources in the room next door with cosparsity, in "ICASSP - IEEE International Conference on Acoustics, Speech, and Signal Processing", Florence, Italy, May 2014 [DOI: 10.1109/ICASSP.2014.6854168], https://hal.inria.fr/hal-00904779.
- [77] S. KITIĆ, N. BERTIN, R. GRIBONVAL.Sparsity and cosparsity for audio declipping: a flexible non-convex approach, in "LVA/ICA 2015 - The 12th International Conference on Latent Variable Analysis and Signal Separation", Liberec, Czech Republic, August 2015, 8, https://hal.inria.fr/hal-01159700.
- [78] M. KOWALSKI, K. SIEDENBURG, M. DÖRFLER. Social Sparsity! Neighborhood Systems Enrich Structured Shrinkage Operators, in "IEEE Transactions on Signal Processing", May 2013, vol. 61, n<sup>o</sup> 10, p. 2498 - 2511 [DOI: 10.1109/TSP.2013.2250967], https://hal.archives-ouvertes.fr/hal-00691774.
- [79] L. LE MAGOAROU, R. GRIBONVAL, A. GRAMFORT. FAµST: speeding up linear transforms for tractable inverse problems, in "European Signal Processing Conference (EUSIPCO)", Nice, France, August 2015, https://hal.archives-ouvertes.fr/hal-01156478.
- [80] L. LE MAGOAROU, R. GRIBONVAL. Chasing butterflies: In search of efficient dictionaries, in "International Conference on Acoustics, Speech and Signal Processing (ICASSP)", Brisbane, Australia, April 2015 [DOI: 10.1109/ICASSP.2015.7178579], https://hal.archives-ouvertes.fr/hal-01104696.
- [81] L. LE MAGOAROU, R. GRIBONVAL. Are There Approximate Fast Fourier Transforms On Graphs?, in "International Conference on Acoustics, Speech and Signal Processing (ICASSP)", Shanghai, China, March 2016, https://hal.archives-ouvertes.fr/hal-01254108.
- [82] L. LE MAGOAROU, R. GRIBONVAL.Flexible Multi-layer Sparse Approximations of Matrices and Applications, in "IEEE Journal of Selected Topics in Signal Processing", June 2016 [DOI: 10.1109/JSTSP.2016.2543461], https://hal.inria.fr/hal-01167948.
- [83] L. LE MAGOAROU. Efficient matrices for signal processing and machine learning, INSA de Rennes, November 2016, https://tel.archives-ouvertes.fr/tel-01412558.
- [84] A. LIUTKUS, R. BADEAU. Generalized Wiener filtering with fractional power spectrograms, in "Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on", IEEE, 2015, p. 266–270.
- [85] C. LOUBOUTIN, F. BIMBOT. Description of Chord Progressions by Minimal Transport Graphs Using the System & Contrast Model, in "ICMC 2016 - 42nd International Computer Music Conference", Utrecht, Netherlands, September 2016, https://hal.archives-ouvertes.fr/hal-01421023.
- [86] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE. Audio-based robot control from interchannel level difference and absolute sound energy, in "IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'16", Daejeon, South Korea, October 2016, p. 1992-1999, https://hal.inria.fr/hal-01355394.
- [87] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE. Binaural auditory interaction without HRTF for humanoid robots: A sensor-based control approach, in "Workshop on Multimodal Sensor-based Control for HRI and soft manipulation, IROS'2016", Daejeon, South Korea, October 2016, https://hal.inria.fr/hal-01408422.

- [88] A. MAGASSOUBA, N. BERTIN, F. CHAUMETTE. First applications of sound-based control on a mobile robot equipped with two microphones, in "IEEE Int. Conf. on Robotics and Automation, ICRA'16", Stockholm, Sweden, May 2016, https://hal.inria.fr/hal-01277589.
- [89] S. NAM, R. GRIBONVAL. Physics-driven structured cosparse modeling for source localization, in "Acoustics, Speech and Signal Processing, IEEE International Conference on (ICASSP 2012)", Kyoto, Japon, IEEE, 2012, http://hal.inria.fr/hal-00659405.
- [90] A. A. NUGRAHA, A. LIUTKUS, E. VINCENT. Multichannel audio source separation with deep neural networks, in "IEEE/ACM Transactions on Audio, Speech, and Language Processing", 2016, vol. 24, n<sup>o</sup> 9, p. 1652–1664.
- [91] A. OZEROV, C. FÉVOTTE.Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation, in "IEEE Transactions on Audio, Speech, and Language Processing", 2010, vol. 18, n<sup>o</sup> 3, p. 550–563.
- [92] G. PUY, P. VANDERGHEYNST, R. GRIBONVAL, Y. WIAUX. Universal and efficient compressed sensing by spread spectrum and application to realistic Fourier imaging techniques, in "EURASIP Journal on Advances in Signal Processing", 2012 [DOI: 10.1186/1687-6180-2012-6], http://hal.inria.fr/inria-00582432.
- [93] Y. SALAÜN, E. VINCENT, N. BERTIN, N. SOUVIRAÀ-LABASTIE, X. JAUREGUIBERRY, D. TRAN, F. BIMBOT.*The Flexible Audio Source Separation Toolbox Version 2.0*, May 2014, ICASSP, https://hal.inria. fr/hal-00957412.
- [94] G. SARGENT. Music structure estimation using multi-criteria analysis and regularity constraints, Université Rennes 1, February 2013, https://tel.archives-ouvertes.fr/tel-00853737.
- [95] A. SCHMIDT, A. DELEFORGE, W. KELLERMANN. Ego-Noise Reduction Using a Motor Data-Guided Multichannel Dictionary, in "International Conference on Intelligent Robots and Systems (IROS), 2016", Daejon, South Korea, IEEE/RSJ, October 2016, p. 1281-1286, https://hal.inria.fr/hal-01415723.
- [96] Y. TRAONMILIN, R. GRIBONVAL.Stable recovery of low-dimensional cones in Hilbert spaces: One RIP to rule them all, in "Applied and Computational Harmonic Analysis", September 2016, https://hal.inria.fr/hal-01207987.

# **Project-Team SERPICO**

Space-timE RePresentation, Imaging and cellular dynamics of molecular COmplexes

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Computational Biology

# **Table of contents**

1.	Personnel	. 813
2.	Overall Objectives	. 814
	2.1. Glossary	814
	2.2. Scientific context and motivations	814
	2.3. Challenges in biological image processing and quantitative microscopy	815
	2.4. Objectives of Serpico in cell imaging	815
	2.5. Organization and collaborations	815
3.	Research Program	. 816
	3.1. Statistics and algorithms for computational microscopy	816
	3.2. From image data to motion descriptors: trajectory computation and dynamics analysis	816
	3.3. Biological and biophysical models and spatial statistics for quantitative bioimaging	817
4.	Application Domains	. 818
	4.1. Modeling and analysis of membrane transport and molecule trafficking at the single cell sc	al <mark>&amp;18</mark>
	4.2. Imaging and analysis of cytokskeleton dynamics during cell migration	819
5.	Highlights of the Year	. 819
6.	New Software and Platforms	. 819
	6.1. ATLAS	819
	6.2. C-CRAFT	820
	6.3. F2D-SAFIR	820
	6.4. GcoPS	820
	6.5. Hullkground	821
	6.6. Motion2D	821
	6.7. ND-SAFIR	822
	6.8. OWF	822
	6.9. QuantEv	822
	6.10. TMA-Lib	823
	6.11. TOTH	823
	6.12. Platforms	823
	6.12.1. Mobyle@Serpico plateform and software distribution	823
	6.12.2. IGRIDA-Serpico cluster	825
7.	New Results	. 825
	7.1. Statistical methods for image denoising and reconstruction	825
	7.2. Algorithms for dejittering and deconvolving fluorescence Tissue MicroArray (TMA) imag	es827
	7.3. Correlation-based method for membrane diffusion estimation during exocytosis in TIRFM	. 827
	7.4. Classification of diffusion dynamics from particle trajectories	829
	7.5. Spatial statistics, point patterns, and colocalization in fluorescence imaging	829
	7.6. Data assimilation and modeling of cell division mechanism	831
	7.7. Quantifying the spatial distribution of intracellular events	831
	7.8. 3D registration for correlative light-electron microscopy	833
	7.9. Fast optical flow methods for 3D fluorescence microscopy	833
	7.10. 3D Convolutional Neural Networks for macromolecule localization in cryo-electron tomogr	ams
	of intact cells	833
	7.11. Estimation of parametric motion models with deep neural networks	834
•	7.12. Motion saliency in video sequences	835
8.	Bilateral Contracts and Grants with Industry	. 835
	8.1. Bilateral Contracts with Industry	835
	8.1.1. Contract with Innopsys: 11ssue microarrays (TMA) image analysis	835
	8.1.2. Contract (CIFRE) with Technicolor: Semantically meaningful motion descriptors for v	deo
	understanding	836

	8.1.3.	Contract with OBSYS: microscope set-up control and inverse problems in microscopy	836						
	8.2. Bila	ateral grants with industry	836						
9.	9. Partnerships and Cooperations								
9.1. Regional Initiatives									
	9.2. National Initiatives								
	9.2.1.	France-BioImaging project	837						
	9.2.2. ANR DALLISH project (2016-2020): Data Assimilation and Lattice LIgh								
	for endocytosis/exocytosis pathway modeling in the whole cell								
9.3. European Initiatives									
	9.4. Inte	rnational Initiatives	838						
	9.4.1.	Informal International Partners	838						
	9.4.2.	Inria Associate Teams Not Involved in an Inria International Labs	838						
	9.5. Inte	rnational Research Visitors	839						
10.	Dissemin	ation	839						
	10.1. Pro	moting Scientific Activities	839						
	10.1.1.	Scientific Events Organisation	839						
	10.1.2.	Scientific Events Selection	839						
	10.1	1.2.1. Member of the Conference Program Committees	839						
	10.1	1.2.2. Reviewer	839						
	10.1.3.	Journal	839						
	10.1.3.1. Member of the Editorial Boards								
	10.1.3.2. Reviewer - Reviewing Activities								
	10.1.4. Invited Talks								
	10.1.5. Leadership within the Scientific Community								
	10.1.6.	Scientific Expertise	840						
	10.1.7.	Research Administration	840						
10.2. Teaching - Supervision - Juries									
	10.2.1. Teaching								
	10.2.2. Supervision								
	10.2.3.	Juries	841						
11.	Bibliogra	aphy	842						

# **Project-Team SERPICO**

*Creation of the Team: 2010 January 01, updated into Project-Team: 2013 July 01* **Keywords:** 

# **Computer Science and Digital Science:**

- A3.1.1. Modeling, representation
- A3.1.2. Data management, quering and storage
- A3.3. Data and knowledge analysis
- A3.4. Machine learning and statistics
- A5.3. Image processing and analysis
- A5.3.2. Sparse modeling and image representation
- A5.3.3. Pattern recognition
- A5.3.4. Registration
- A5.4.4. 3D and spatio-temporal reconstruction
- A5.4.5. Object tracking and motion analysis
- A5.4.6. Object localization
- A5.9.1. Sampling, acquisition
- A5.9.2. Estimation, modeling
- A5.9.3. Reconstruction, enhancement
- A5.9.6. Optimization tools
- A6.1.2. Stochastic Modeling (SPDE, SDE)
- A6.1.3. Discrete Modeling (multi-agent, people centered)
- A6.1.4. Multiscale modeling
- A6.1.5. Multiphysics modeling
- A6.2.3. Probabilistic methods
- A6.2.4. Statistical methods
- A6.2.6. Optimization
- A6.3.1. Inverse problems
- A6.3.2. Data assimilation
- A6.3.3. Data processing

## **Other Research Topics and Application Domains:**

- B1.1.1. Structural biology
- B1.1.3. Cellular biology
- B1.1.9. Bioinformatics
- B1.1.10. Mathematical biology
- B2.2.3. Cancer
- B2.6. Biological and medical imaging

# **1. Personnel**

#### **Research Scientists**

Charles Kervrann [Team leader, Inria, Senior Researcher, HDR]

Patrick Bouthemy [Inria, Senior Researcher, HDR] Sylvain Prima [Inria, Researcher, Visiting scientist]

#### **Faculty Member**

Frédéric Lavancier [Univ de Nantes, Associate Professor, until Aug 2017, HDR]

#### **Technical Staff**

Giovanni Petrazzuoli [Inria, granted by OBSYS, until Jun 2017]

#### **PhD Students**

Vincent Briane [Inria, granted by Inria & ENSAI-CREST, until Nov 2017] Anca-Georgiana Caranfil [Univ de Rennes 1, MESR grant] Yunjiao Lu [INRA, INRA grant, from Oct 2017] Léo Maczyta [Inria, granted by DGA & Région Bretagne, from Oct 2017] Sandeep Manandhar [Inria, CORDI-S grant] Emmanuel Moebel [Inria, granted by Fourmentin-Guilbert & Région Bretagne] Juan Manuel Perez Rua [Technicolor, granted by CIFRE, until Jan 2018] Antoine Salomon [Inria, granted by ANR DALLISH, from Nov 2017] Bertha Mayela Toledo Acosta [granted by Conacyt, Mexico]

#### Administrative Assistant

Huguette Bechu [Inria]

# 2. Overall Objectives

## 2.1. Glossary

- FLIM (Fluorescence Lifetime Microscopy Imaging): imaging of fluorescent molecule lifetimes.
- **PALM** (Photo-Activated Localization Microscopy): high-resolution microscopy using stochastic photo-activation of fluorophores and adjustment of point spread functions [38].
- **SIM** (Structured Illumination Microscopy): high-resolution light microscopy using structured patterns and interference analysis [42].
- **TIRF** (Total Internal Reflectance): 2D optical microscopy using evanescent waves and total reflectance [37].
- **Cryo-EM** (Cryo-Electron Tomography): 3D representation of sub-cellular and molecular objects of 5-20 nanometres, frozen at very low temperatures, from 2D projections using a transmission electron microscope.

# 2.2. Scientific context and motivations

During the past two decades, biological imaging has undergone a revolution in the development of new microscopy techniques that allow visualization of tissues, cells, proteins and macromolecular structures at all levels of resolution, physiological states, chemical composition and dynamics. Thanks to recent advances in optics, digital sensors and labeling probes (e.g., Colored Fluorescence Protein), one can now visualize sub-cellular components and organelles at the scale of several hundreds of nanometers to a few dozens nanometers. As a result, fluorescent microscopy and multimodal imaging (fluorophores at various wavelengths) have become the workhorse of modern biology. All the technological advances in microscopy have created new issues and challenges for researchers in quantitative image processing and analysis. Since the digital processing is now part of the imaging loop, image processing may even drive imaging. A brilliant example of this shift in paradigm is super-resolution localization microscopy (PALM, STED), which was awarded the 2014 Nobel Prize in Chemistry.

# 2.3. Challenges in biological image processing and quantitative microscopy

In most cases, modern microscopy in biology is characterized by a large number of dimensions that fit perfectly with the complexity of biological features: two or three spatial dimensions, at macro to nano-scales, and one temporal dimension, sometimes spectrally defined and often corresponding to one particular biomolecular species. Dynamic microscopy is also characterized by the nature of the observable objects (cells, organelles, single molecules, ...), by the large number of small size and mobile elements (chromosomes, vesicles, ...), by the complexity of the dynamic processes involving many entities or group of entities sometimes interacting, by particular phenomena of coalescence often linked to image resolution problems, finally by the association, dissociation, recomposition or constitution of those entities (such as membrane fusion and budding). Thus, the corpus of data to be considered for any analysis involving multiple image series acquisitions is massive (up to few GigaBytes per hour). Therefore, it becomes necessary to facilitate and rationalize the production of those multidimensional data, to improve post acquisition analysis, and to favor the organization and the interpretation of the information extracted from this data corpus. It motivates innovative methods and concepts for data fusion, image registration, super-resolution, data mining... More importantly, modern microscopy has led to recent breakthroughs, related to the potential interactions between molecules in the cell. A longterm research consists now in inferring the relationships between the dynamics of macromolecules and their functions. Research on computational biology and quantitative bioimaging lies at the core of the activities of Serpico team.

# 2.4. Objectives of Serpico in cell imaging

In order to tackle the aforementioned challenges, the Serpico team aims to develop innovative approaches and paradigms for image reconstruction, 3D molecule tracking and motion estimation, and biophysical parameter estimation to face the huge data volumes acquired with cutting-edge microscopy set-ups. To this end, applied mathematics, image processing and analysis have to be considered in association with biophysics and biology. To be successful, a sustained synergy between all these scientific domains is necessary. To improve state-of-the-art methods and solve important problems in computational bioimaging, the members of Serpico especially address the following topics:

- Image restoration/reconstruction motivated by preserving cell integrity (photo-toxicity versus exposure time) and image analysis in multidimensional microscopy;
- Motion analysis and computation of molecule trajectories in live-cell imaging to study molecular interactions in space and time;
- Computational simulation, modelling and estimation of molecule trafficking and interactions at different spatial and temporal scales.

The resulting mathematical models and algorithms will help biologists to decipher molecular processes in fundamental biology and will be exploited for health applications: disease diagnosis, detection of genomic instabilities, deterioration of cell cycle, cancer prevention.

We have successfully developed statistical and variational aggregation methods for image denoising and optical flow, and elaborated powerful methods for image colocalization, diffusion estimation, trajectory estimation-classification, and multimodal registration. An additional issue was the design and distribution of software tools for the biological image analysis and microscopy communities. Finally, the team has focused on the cellular and molecular mechanisms involved in molecule and protein transport and trafficking at the scale of a single cell. Our contributions are detailed in the next sections along three research axes.

## 2.5. Organization and collaborations

In collaboration with UMR 144 CNRS-Institut Curie ("Space Time imaging of Endomembranes and organelles Dynamics" team) and PICT-IBiSA (Cell and Tissue Imaging Facilities), the members of the Serpico team have participated in several projects (PhD and post-doc supervision, contracts...) in the field of cell biology and microscopy. We have promoted non-parametric methods since prior knowledge cannot be easily taken

into account for extracting unattended but desired information from image data. We have also proposed user-friendly algorithms for processing 2D and 3D image sequences. The projects of Serpico were in line with several studies led in the UMR 144 CNRS-Institut Curie Unit. A subset of studies was related to instrumentation in electronic and photonic microscopy (PICT-IBiSA platform) including computational aspects on the reconstruction and enhancement of images related to sub-diffraction light microscopy and multimodal approaches. Serpico projects relied partially on the advances of these instrumental projects and a positive synergy was established.

# **3. Research Program**

### 3.1. Statistics and algorithms for computational microscopy

Fluorescence microscopy limitations are due to the optical aberrations, the resolution of the microscopy system, and the photon budget available for the biological specimen. Hence, new concepts have been defined to address challenging image restoration and molecule detection problems while preserving the integrity of samples. Accordingly, the main stream regarding denoising, deconvolution, registration and detection algorithms advocates appropriate signal processing framework to improve spatial resolution, while at the same time pushing the illumination to extreme low levels in order to limit photo-damages and phototoxicity [7], [6]. As a consequence, the question of adapting cutting-edge signal denoising and deconvolution, object detection, and image registration methods to 3D fluorescence microscopy imaging has retained the attention of several teams over the world.

In this area, the Serpico team has developed a strong expertise in key topics in computational imaging including image denoising and deconvolution, object detection and multimodal image registration. Several algorithms proposed by the team outperformed the state-of-the-art results, and some developments are compatible with "high-throughput microscopy" and the processing of several hundreds of cells. We especially promoted non local, non-parametric and patch-based methods to solve well-known inverse problems or more original reconstruction problems. A recent research direction consists in adapting the deep learning concept to solve challenging detection and reconstruction problems in microscopy. We have investigated convolution neural networks to detect small macromolecules in 3D noisy electron images with promising results. The next step consists in proposing smart paradigms and architectures to save memory and computations.

More generally, many inverse problems and image processing become intractable with modern 3D microscopy, because very large temporal series of volumes (200 to 1000 images per second for one 3D stack) are acquired for several hours. Novel strategies are needed for 3D image denoising, deconvolution and reconstruction since computation is extremely heavy. Accordingly, we will adapt the estimator aggregation approach developed for optical flow computation to meet the requirements of 3D image processing. We plan to investigate regularization-based aggregation energy over super-voxels to reduce complexity, combined to modern optimization algorithms. Finally, we will design parallelized algorithms that fast process 3D images, perform energy minimization in few seconds per image, and run on low-cost graphics processor boards (GPU).

# **3.2.** From image data to motion descriptors: trajectory computation and dynamics analysis

Several particle tracking methods for intracellular analysis have been tailored to cope with different types of cellular and subcellular motion down to Brownian single molecule behavior. Many algorithms were carefully evaluated on the particle tracking challenge dataset published in the Nature Methods journal in 2014 [8]. Actually, there is no definitive solution to the particle tracking problem which remains application-dependent in most cases. The work of Serpico in particle motion analysis is significant in multiple ways, and inserts within a very active international context. One of the remaining key open issues is the tracking of objects with heterogeneous movements in crowded configurations. Moreover, particle tracking methods are not always adapted for motion analysis, especially when the density of moving features hampers the individual extraction

of objects of interest undergoing complex motion. Estimating flow fields can be more appropriate to capture the complex dynamics observed in biological sequences. The existing optical flow methods can be classified into two main categories: i/ local methods impose a parametric motion model (e.g. local translation) in a given neighborhood; ii/ global methods estimate the dense motion field by minimizing a global energy functional composed of a data term and a regularization term.

The Serpico team has developed a strong expertise in key topics, especially in object tracking for fluorescence microscopy, optical flow computation and high-level analysis of motion descriptors and trajectories. Several algorithms proposed by the team are very competitive when compared to the state-of-the-art results, and our new paradigms offer promising ways for molecule traffic quantification and analysis. Amongst the problems that we currently address, we can mention: computation of 3D optical flow for large-size images, combination of two frame-based differential methods and sparse sets of trajectories, detection and analysis of unexpected local motion patterns in global coherent collective motion. Development of efficient numerical schemes will be central in the future but visualization methods are also crucial for evaluation and quality assessment. Another direction of research consists in exploiting deep learning to 3D optical flow so as to develop efficient numerical schemes that naturally capture complex motion patterns. Investigation in machine learning and statistics will be actually conducted in the team in the two first research axes to address a large range of inverse problems in bioimaging. Deep learning is an appealing approach since expertise of biologists, via iterative annotation of training data, will be included in the design of image analysis schemes.

# **3.3.** Biological and biophysical models and spatial statistics for quantitative bioimaging

A number of stochastic mathematical models were proposed to describe various intracellular trafficking, where molecules and proteins are transported to their destinations via free diffusion, subdiffusion and ballistic motion representing movements along the cytoskeleton networks assisted by molecular motors. Accordingly, the study of diffusion and stochastic dynamics has known a growing interest in bio-mathematics, biophysics and cell biology with the popularization of fluorescence dynamical microscopy and super-resolution imaging. In this area, the competing teams mainly studied MSD and fluorescence correlation spectroscopy methods.

In the recent period, the Serpico team achieved important results for diffusion-related dynamics involved in exocytosis mechanisms. Robustness to noise has been well investigated, but robustness to environmental effects has yet to be effectively achieved. Particular attention has been given to the estimation of particle motion regime changes, but the available results are still limited for analysing short tracks. The analysis of spatiotemporal molecular interactions from set of 3D computed trajectories or motion vector fields (e.g., co-alignment) must be investigated to fully quantify specific molecular machineries. We have already made efforts in that directions this year (e.g., for colocalization) but important experiments are required to make our preliminary algorithms reliable enough and well adapted to specific transport mechanisms.

Accordingly, we will study quantification methods to represent interactions between molecules and trafficking around three lines of research. First, we will focus on 3D space-time global and local object-based coorientation and co-alignment methods, in the line of previous work on colocalization, to quantify interactions between molecular species. In addition, given N tracks associated to N molecular species, interaction descriptors, dynamics models and stochastic graphical models representing molecular machines will be studied in the statistical data assimilation framework. Second, we will analyse approaches to estimate molecular mobility, active transport and motion regime changes from computed trajectories in the Lagrangian and Eulerian settings. We will focus on the concept of super-resolution to provide spatially high-resolved maps of diffusion and active transport parameters based on stochastic biophysical models and sparse image representation. Third, we plan to extend the aggregation framework dedicated to optical flow to the problem of diffusion-transport estimation. Finally, we will investigate data assimilation methods to better combine algorithms, models, and experiments in an iterative and virtuous circle. The overview of ultrastructural organization will be achieved by additional 3D electron microscopy technologies.

# 4. Application Domains

4.1. Modeling and analysis of membrane transport and molecule trafficking at the single cell scale



Figure 1. Cargo Langerin Trafficking controlled by Rab11A/Rab11FIP2/MyoVb platform.

In the past recent years, research carried at UMR 144 CNRS-Institut Curie ("Space Time imaging of Endomembranes and organelles Dynamics" team) contributed to a better understanding of the intracellular compartimentation of specialized model cells such as melanocytes and Langerhans cells, the components and structural events involved in the biogenesis of their specialized organelles: melanosomes and Birbeck granules, respectively. These studies have started to highlight: i/ multiple sorting and structural events involved in the biogenesis; ii/ complexity of the endo-melanosomal network of these highly specialized cells; iii/ complex molecular architecture organizing and coordinating their dynamics; iv/ intracellular transport steps affected in genetic diseases, among which the Hermansky Pudlak syndrome (HPS) or involved in viral infection (HIV and Langerin in Langerhans cells).

In this context, the central aim of SERPICO is to understand how the different machineries of molecular components involved are interconnected and coordinated to generate such specialized structures. We need to address the following topics:

- 1. developing new bioimaging approaches to observe and statistically analyze such coordinated dynamics in live material;
- 2. correlating this statistically relevant spatiotemporal organization of protein networks with the biological architectures and at the ultrastructural level;
- 3. modeling intracellular transport of those reference biological complex systems and proposing new experimental plans in an iterative and virtuous circle;
- 4. managing and analyzing the workflow of image data obtained along different multidimensional microscopy modalities.

These studies are essential to unravel the complexity of the endomembrane system and how different machineries evolve together (e.g. see Fig. 1). They help to control cell organization and function at different scales through an integrative workflow of methodological and technological developments.

At long term, these studies will shed light on the cellular and molecular mechanisms underlying antigen presentation, viral infection or defense mechanisms, skin pigmentation, the pathogenesis of hereditary genetic disorders (lysosomal diseases, immune disorders) and on the mechanisms underlying cell transformation. Our methodological goal is also to link dynamics information obtained through diffraction limited light microscopy, eventually at a time regime compatible with live cell imaging. The overview of ultrastructural organization will be achieved by complementary electron microscopical methods. Image visualization and quantitative analysis are of course important and essential issues in this context.

## 4.2. Imaging and analysis of cytokskeleton dynamics during cell migration

The ability to migrate in space is among the most fundamental functions of eukaryotic cells and thus is one of the best-studied phenomena in biology. During embryonic development, cell movements result in a massive reorganization of the embryo, from a simple spherical ball of cells into a multi-layered organism; many of the cells at or near the surface of the embryo move to a new, more interior location. Moreover, inadequate or inappropriate migration of immune cells is also critically important for the delivery of protective immune responses to tissues and for wound healing. Finally, cell migration may facilitate the dissemination of tumor cells in blood and organs and eventually the formation of secondary tumors and metastases.

It has been established that the cytoskeleton, composed of actin filaments, microtubules and intermediate filaments (elongated structures with a diameter of a few dozens of nanometers), is essential for several cell mechanisms, including cell migration, cell division and molecule trafficking:

i/ the actin filaments promote cell protrusion, adhesion and retraction;

ii/ the microtubules are the support of molecule traffic and cell polarization;

iii/ the intermediate filaments are hypothesized to control microtubule organization.

Nevertheless, the mechanical and chemical states of migrating cells under various external conditions remain largely unknown. In the last decade, high-resolution microscopy methods led to the discovery of novel aspects of cell migration. Most approaches and models are limited to migration in 2D, justified by the flatness of the cell-motile mechanisms. However, the mechanical patterns that govern migration in 2D models are often not essential for efficient migration in 3D. Accordingly, recent very challenging 3D models of cells moving on flat surfaces have begun to emerge. The key challenge, however, is to understand how a 3D motile cell crawls through the 3D extracellular matrix.

The objective of SERPICO is to develop high-end signal processing and computer vision tools to unfold the dynamical coordination of microtubules, actin filaments and intermediate filaments in 3D, involved in cell migration, cell division and molecule trafficking.

# 5. Highlights of the Year

# 5.1. Highlights of the Year

The Serpico team will be the organizer of the 7th International Conference on "Quantitative BioImaging" (QBI) in January 2019 (300 attendees) in Rennes.

Juan Manual Perez Ru, Vincent Briane and Hoai-Nam Nguyeen defended their PhD thesis in December 2017.

# 6. New Software and Platforms

# 6.1. ATLAS

KEYWORDS: Image segmentation - Object detection - Photonic imaging - Image analysis - Fluorescence microscopy

FUNCTIONAL DESCRIPTION: The ATLAS software enables to detect spots in 2D fluorescence images. The spot size is automatically selected and the detection threshold adapts to the local image contrasts. ATLAS relies on the Laplacian of Gaussian (LoG) filter, which both reduces noise and enhances spots. A multiscale representation of the image is built to automatically select the optimal LoG variance. Local statistics of the LoG image are estimated in a Gaussian window, and the detection threshold is pointwise inferred from a probability of false alarm (PFA). The user only has to specify: i/ size of the Gaussian window, ii/ PFA value. The Gaussian window must be about the size of the background structures, increasing the PFA increases the number of detections.

- Participants: Patrick Bouthemy, Charles Kervrann, Jean Salamero, Jérôme Boulanger and Antoine Basset
- Partner: UMR 144 CNRS Institut Curie
- Contact: Patrick Bouthemy
- Publication: Adaptive spot detection with optimal scale selection in fluorescence microscopy images
- URL: http://mobyle-serpico.rennes.inria.fr/cgi-bin/portal.py#forms::ATLAS

# 6.2. C-CRAFT

KEYWORDS: Fluorescence microscopy - Photonic imaging - Image analysis - Detection - 3D - Health - Biology - Segmentation

FUNCTIONAL DESCRIPTION: The C-CRAFT software enables to jointly segment small particles and estimate background in 2D or 3D fluorescence microscopy image sequences. The vesicle segmentation and background estimation problem is formulated as a global energy minimization problem in the Conditional Random Field framework. A patch-based image representation is used to detect spatial irregularity in the image. An iterative scheme based on graph-cut algorithm is proposed for energy minimization.

- Participants: Patrick Bouthemy, Jean Salamero, Charles Kervrann and Thierry Pécot
- Partner: UMR 144 CNRS Institut Curie
- Contact: Charles Kervrann
- Publication: Background Fluorescence Estimation and Vesicle Segmentation in Live Cell Imaging with Conditional Random Fields
- URL: http://mobyle-serpico.rennes.inria.fr/cgi-bin/portal.py#forms::C-CRAFT

## 6.3. F2D-SAFIR

KEYWORDS: Biomedical imaging - Photonic imaging - Fluorescence microscopy - Image processing FUNCTIONAL DESCRIPTION: The F2D-SAFIR software removes mixed Gaussian-Poisson noise in large 2D images, typically 10000 x 10000 pixels, in a few seconds. The method is unsupervised and is a simplified version of the method related to the ND-SAFIR software. The software is dedicated to microarrays image denoising for disease diagnosis and multiple applications (gene expression, genotyping, aCGH, ChIP-chip, microRNA, ...).

- Participant: Charles Kervrann
- Partner: INRA
- Contact: Charles Kervrann

# 6.4. GcoPS

KEYWORDS: Photonic imaging - Fluorescence microscopy - Image processing - Statistic analysis

FUNCTIONAL DESCRIPTION: The GCOPS (Geo-Co-Positioning System) software is dedicated to the colocalization of fluorescence image pairs for both conventional and super-resolution microscopy. The procedure is only controlled by a p-value and tests whether the Pearson correlation between two binary images is significantly positive. It amounts to quantifying the interaction strength by the area/volume of the intersection between the two binary images viewed as random distributions of geometrical objects. Under mild assumptions, it turns out that the appropriately normalized Pearson correlation follows a standard normal distribution under the null hypothesis if the number of image pixels is large. Unlike previous methods, GcoPS handles 2D and 3D images, variable SNRs and any kind of cell shapes. It is able to co-localize large regions with small dots, as it is the case in TIRF-PALM experiments and to detect negative co-localization. The typical processing time is two milliseconds per image pair in 2D and a few seconds in 3D, with no dependence on the number of objects per image. In addition, the method provides maps to geo-co-localize molecule interactions in specific image regions.

- Participants: Frédéric Lavancier, Thierry Pécot and Liu Zengzhen
- Partners: Université de Nantes UMR 144 CNRS Institut Curie
- Contact: Charles Kervrann
- Publication: A Fast Automatic Colocalization Method for 3D Live Cell and Super-Resolution Microscopy
- URL: http://icy.bioimageanalysis.org/plugin/GcoPS

## 6.5. Hullkground

KEYWORDS: Biomedical imaging - Photonic imaging - Fluorescence microscopy - Image processing FUNCTIONAL DESCRIPTION: The HullkGround software decomposes a fluorescence microscopy image sequence into two dynamic components: i) an image sequence showing mobile objects, ii) an image sequence showing the slightly moving background. Each temporal signal of the sequence is processed individually and analyzed with computational geometry tools. The convex hull is estimated automatically for each pixel and subtracted to the original signal. The method is unsupervised, requires no parameter tuning and is a simplified version of the shapes-based scale-space method.

- Participants: Anatole Chessel, Charles Kervrann and Jean Salamero
- Partner: UMR 144 CNRS Institut Curie
- Contact: Charles Kervrann
- URL: http://mobyle-serpico.rennes.inria.fr/cgi-bin/portal.py#forms::Hullkground

# 6.6. Motion2D

KEYWORDS: Image sequence - Motion model - 2D

FUNCTIONAL DESCRIPTION: The Motion2D software is a multi-platform object-oriented library to estimate 2D parametric motion models in an image sequence. It can handle several types of motion models, namely, constant (translation), affine, and quadratic models. Moreover, it includes the possibility of accounting for a global variation of illumination and more recently for temporal image intensity decay (e.g. due to photobleaching decay in fluorescence microscopy). The use of such motion models has been proved adequate and efficient for solving problems such as optic flow computation, motion segmentation, detection of independent moving objects, object tracking, or camera motion estimation, and in numerous application domains (video surveillance, visual servoing for robots, video coding, video indexing), including biological imaging (image stack registration, motion compensation in videomicroscopy). Motion2D is an extended and optimized implementation of the robust, multi-resolution and incremental estimation method (exploiting only the spatiotemporal derivatives of the image intensity function). Real-time processing is achievable for motion models involving up to six parameters. Motion2D can be applied to the entire image or to any pre-defined window or region in the image.

RELEASE FUNCTIONAL DESCRIPTION: Modifications and improvements in the PNG image file support. Support RAW and Mpeg2 video format as input (see CReader). The available video format which can be handled by the motion estimator are given by CReader::EReaderFormat. For the results, video sequences can be writen using the format specified by CWriter::EWriterFormat. Support Fedora 3 (g++ 3.4.2).

- Participants: Charles Kervrann, Fabien Spindler, Jean Marc Odobez, Patrick Bouthemy and Thierry Pécot
- Contact: Patrick Bouthemy
- URL: http://www.irisa.fr/vista/Motion2D/

# 6.7. ND-SAFIR

KEYWORDS: Fluorescence microscopy - Photonic imaging - Image analysis - Health - Biomedical imaging SCIENTIFIC DESCRIPTION: ND-SAFIR is a software for denoising n-dimentionnal images especially dedicated to microscopy image sequence analysis. It is able to deal with 2D, 3D, 2D+time, 3D+time images have one or more color channel. It is adapted to Gaussian and Poisson-Gaussian noise which are usually encountered in photonic imaging. Several papers describe the detail of the method used in ndsafir to recover noise free images (see references).

- Participants: Charles Kervrann, Patrick Bouthemy, Jean Salamero and Jérôme Boulanger
- Partners: INRA PiCT UMR 144 CNRS Institut Curie
- Contact: Charles Kervrann
- URL: http://serpico.rennes.inria.fr/doku.php?id=software:nd-safir:index

# 6.8. OWF

KEYWORDS: Image filter - Image processing - Statistics

FUNCTIONAL DESCRIPTION: The OWF software enables to denoise images corrupted by additive white Gaussian noise. In the line of work of the Non-Local means and ND-SAFIR algorithms, this adaptive estimator is based on the weighted average of observations taken in a neighborhood with weights depending on the similarity of local patches. The idea is to compute adaptive weights that best minimize an upper bound of the pointwise L2 risk. The spatially varying smoothing parameter is automatically adjusted to the image context. The proposed algorithm is fast and easy to control and is competitive when compared to the more sophisticated NL-means filters.

- Participants: Ion Grama, Quansheng Liu and Qiyu Jin
- Partner: University of Bretagne-Sud
- Contact: Charles Kervrann
- Publication: Non-local means and optimal weights for noise removal
- URL: http://serpico.rennes.inria.fr/doku.php?id=software:owf

## 6.9. QuantEv

KEYWORDS: Photonic imaging - Fluorescence microscopy - Biomedical imaging - Image analysis - Image sequence - Statistic analysis

FUNCTIONAL DESCRIPTION: The QUANTEV software analyzes the spatial distribution of intracellular events represented by any static or dynamical descriptor, provided that the descriptors are associated with spatial coordinates. QUANTEV first computes 3D histograms of descriptors in a cylindrical coordinate system with computational cell shape normalization, enabling comparisons between cells of different shape. Densities are obtained via adaptive kernel density estimation, and we use the Circular Earth Mover's Distance to measure the dissimilarity between densities associated to different experimental conditions. A statistical analysis on these distances reliably takes into account the biological variability over replicated experiments.

- Participants: Jean Salamero, Jérôme Boulanger and Liu Zengzhen
- Partner: UMR 144 CNRS Institut Curie
- Contact: Charles Kervrann
- Publication: QuantEv: quantifying the spatial distribution of intracellular events
- URL: http://mobyle-serpico.rennes.inria.fr/cgi-bin/portal.py#forms::QuantEv-Densities

# 6.10. TMA-Lib

KEYWORDS: Photonic imaging - Fluorescence microscopy - Biomedical imaging - Image processing FUNCTIONAL DESCRIPTION: The TMA-LIB enables to jointly detect suing adaptive wavelet transform, segment with parametric active contours and restore (i.e., artifact correction and deconvolution) TMA (Tissue MicroArrays) images.

- Participants: Cyril Cauchois, Vincent Paveau and Hoai Nam Nguyen
- Partner: Innopsys
- Contact: Charles Kervrann
- Publications: A variational method for dejittering large fluorescence line scanner images Generalized Sparse Variation Regularization for Large Fluorescence Image Deconvolution - ATMAD : robust image analysis for Automatic Tissue MicroArray De-arraying

# 6.11. TOTH

KEYWORDS: Photonic imaging - Fluorescence microscopy - Biomedical imaging - Classification - Statistical categorisation techniques - Statistics - Image sequence - Visual tracking

FUNCTIONAL DESCRIPTION: The TOTH software classifies trajectories of biomolecules computed with tracking algorithms. Trajectories in living cells are generally modelled with three types of diffusion processes: (i) free diffusion, (ii) subdiffusion or (iii) superdiffusion. We used a test approach with the Brownian motion as the null hypothesis, and developed a non-parametric three-decision test whose alternatives are subdiffusion and superdiffusion. First, we built a single test procedure for testing a single trajectory. Second, we proposed a multiple test procedure for testing a collection of trajectories. These procedures control respectively the type I error and the false discovery rate. Our approach can be considered as an alternative to the Mean Square Displacement (MSD) method commonly used to address this issue. It gives more reliable results as confirmed by our Monte Carlo simulations and evaluations on real sequences of images depicting protein dynamics acquired with TIRF or SPT-PALM microscopy.

- Participants: Vincent Briane and Myriam Vimond
- Partner: ENSAI
- Contact: Charles Kervrann
- Publication: A Statistical Analysis of Particle Trajectories in Living Cells
- URL: http://serpico.rennes.inria.fr/doku.php?id=software:thot:index

## 6.12. Platforms

#### 6.12.1. Mobyle@Serpico plateform and software distribution

Mobyle@SERPICO							(guest) set email   sign-in   activate   sign-out refresh workspace
SERPICO team (INRIA Rennes - Bretage	me Atlantiqu	ue) is partner of Welcome	France-Biol	maging Data Bookmarks	Jobs	Tutorials	
Programs Backwarping • CRFMovingSpotDetection • Hullkground • KLTracker • Motion2D • Motion2D • MS-Detect • NDSafr		Welcome to Mobyle, a portal for bioinformatics analyses Space timE RePresentation, Imaging and cellular dynamics of molecular					
OpticalFlow     OpticalFlowStack  Tutorials     registration		COmplexes Programs available  Backwarping: Warp sequence with parametric motion model CRHwoingSpotDetection: Detecting moving spots/vesicles using Conditional Random Fields CRHwoingSpotDetection: Robust detection of fluorescence accumulation over time in video-microscopy HullkGround: Separation of moving and non moving part in a sequences KLTracker: Track vesicle and POI in image sequences Motion2D: Estimate 2D parametric motion model KLTracker: Track vesicle and POI in image sequences by background subtraction ND-SAFIR: Denoise N-Dimensional images Optical-flow: Compute Optical Flow between 2 images Optical-Flow: Compute Optical Flow between each pair of images in a TIFF stack					
		Credits Mobyle is a p More informa	latform deve tion about th	loped jointly by the Ins is project can be found	titut Pasteu here. BIOİMAGINO	Biology IT Cer	nter and the Ressource Parisienne en Bloinformatique Structurale.

Figure 2. Mobyle@SERPICO web portal.

The objective is to disseminate the distribution of SERPICO image processing software in the community of cell biology and cell imaging.

**Free binaries:** software packages have been compiled for the main operating systems (Linux, MacOS, Windows) using CMake (see http://www.cmake.org/). They are freely available on the team website under a proprietary license (e.g. ND-SAFIR and HULLKGROUND are distributed this way at http://serpico.rennes. inria.fr/doku.php?id=software:index).

**Mobyle@Serpico web portal**: An on-line version (http://mobyle-serpico.rennes.inria.fr) of the image processing algorithms has been developped using the Mobyle framework (Institut Pasteur, see http://mobyle. pasteur.fr/). The main role of this web portal (see Fig. 2) is to demonstrate the performance of the programs developed by the team: QUANTEV, C-CRAFT[14], ATLAS[1], HULLKGROUND[40], KLTRACKER[44], MO-TION2D[43], MS-DETECT[41], ND-SAFIR[6], OPTICALFLOW and FLUX ESTIMATION [14]. The web interface makes our image processing methods available for biologists at Mobyle@SERPICO (http://mobyle-serpico.rennes.inria.fr/cgi-bin/portal.py#welcome) without any installation or configuration on their own. The size of submitted images is limited to 200 MegaBytes per user and all the results are kept 15 days. The web portal and calculations run on a server with 2 CPU x 8 cores, 64 GigaBytes of RAM (500 MegaBytes for each user / Data is saved for 3 months).

**ImageJ plugins**: IMAGEJ (see http://rsb.info.nih.gov/ij/) is a widely used image visualization and analysis software for biologist users. We have developed IMAGEJ plug-in JAVA versions of the following software: ND-SAFIR [6], HULLKGROUND [40], MOTION2D [43], ATLAS [1]. The C-CRAFT algorithm [14] has been developed for the image processing ICY platform (http://icy.bioimageanalysis.org/).

- Contact: Charles Kervrann, Charles Deltel (Inria Rennes SED).
- Partner: UMR 144 CNRS-Institut Curie and France-BioImaging.

#### 6.12.2. IGRIDA-Serpico cluster

The IGRIDA-Serpico cluster of 200 nodes is opened for end-users for large scale computing and data sets processing (200 TeraBytes).

- Batch Scheduler: OAR
- File management: Puppet / Git / Capistrano
- **OS:** Linux Debian 7
- User connexion: public ssh key
- Contact: Charles Kervrann, Charles Deltel (Inria Rennes SED).

# 7. New Results

## 7.1. Statistical methods for image denoising and reconstruction

Participants: Emmanuel Moebel, Charles Kervrann.

In the line of the Non-Local (NL) means [39] and ND-SAFIR [11], [12], [6] denoising algorithms, we have proposed a novel adaptive estimator based on the weighted average of observations taken in a neighborhood with weights depending on image data. The idea is to compute adaptive weights that best minimize an upper bound of the pointwise  $L_2$  risk. In the framework of adaptive estimation, we show that the "oracle" weights depend on the unknown image and are optimal if we consider triangular kernels instead of the commonly-used Gaussian kernel. Furthermore, we propose a way to automatically choose the spatially varying smoothing parameter for adaptive denoising. Under conventional minimal regularity conditions, the obtained estimator converges at the usual optimal rate. The implementation of the proposed algorithm is also straightforward. The simulations show first that our algorithm improves significantly the classical NL-means. Second, the simulations demonstrate that it is competitive when compared to state-of-the-art denoisers both in terms of PSNR values and visual quality.

Meanwhile, we investigated statistical aggregation methods which optimally combine several estimators to produce a boosted solution [13]. This approach has been especially investigated to restore spectral information in the missing wedge (MW) in cryo-electron tomography (CET). The MW is known to be responsible for several types of imaging artifacts, and arises because of limited angle tomography: it is observable in the Fourier domain and is depicted by a region where Fourier coefficient values are unknown (see Fig. 3). The proposed stochastic method tackles the restoration problem by filling up the MW by iterating following steps: adding noise into the MW (step 1) and applying a denoising algorithm (step 2). The role of the first step is to propose candidates for the missing Fourier coefficients and the second step acts as a regularizer. A constraint is added in the spectral domain by imposing the known Fourier coefficients to be unchanged through iterations. Different denoising algorithms (BM3D, NL-Bayes, NL-means...) have been compared. Furthermore, different transforms have been tested in order to apply the constraint (Fourier transform, Cosine transform, pseudopolar Fourier transform). Finally, we showed that this strategy can be embedded into a Monte-Carlo simulation framework and amounts to computing an aggregated estimator [13]. Convincing results have been achieved (see Fig. 3) using the Fourier Shell Correlation (FSC) as an evaluation metric.

#### References: [18]

Collaborators: Qiyu Jin (School of Mathematical Science, Inner Mongolia University, China),

Ion Grama and Quansheng Liu (University of Bretagne-Sud, Vannes),

Damien Larivière (Fondation Fourmentin-Guilbert),



Figure 3. Experimental sub-tomogram containing ribosomes attached to a membrane. (a) Top row: original data in the spectral (left) and spatial (middle) domains and 3D view of the thresholded data (right). Bottom row: denoised data shown as above. (b) FSC and constrained FSC measures of the method input (in black) and output (in red). All measures are wrt the same reference.

Julio Ortiz (Max-Planck Institute, Martinsried, Germany).

# 7.2. Algorithms for dejittering and deconvolving fluorescence Tissue MicroArray (TMA) images

Participant: Charles Kervrann.

In the thesis of H.-N. Nguyen, we developed dedicated image processing methods to improve quality of Tissue Microarray (TMA) images acquired by fluorescence scanners. Images are first acquired pixel by pixel along each line, with a change of scan direction between two subsequent lines. Such scanning system often suffers from pixel mis-positioning (jitter) due to imperfect synchronization of mechanical and electronic components. To correct these scanning artifacts, we proposed a variational method based on the estimation of pixel displacements on subsequent lines. This method, inspired from optical flow methods, consists in estimating a dense displacement field by minimizing an energy function composed of a non-convex data fidelity term and a convex regularization term. We used half-quadratic splitting technique to decouple the original problem into two small sub-problems: one is convex and can be solved by standard optimization algorithm, the other is non-convex but can be solved by a complete search. We showed that our method is able to remove efficiently the rolling effect due to jitter, even in the case of huge images and large non-integer displacements.

Second, to improve the resolution of acquired fluorescence images, we introduced a method of image deconvolution by considering a family of convex regularizers. The considered regularizers are generalized from the concept of Sparse Variation which combines the L1 norm and Total Variation (TV) to favors the colocalization of high-intensity pixels and high-magnitude gradient. The experiments showed that the proposed regularization approach produces competitive deconvolution results on fluorescence images, compared to those obtained with other approaches such as TV or the Schatten norm of Hessian matrix. The final deconvolution algorithm has been dedicated to large 2D 20000 × 60000 images acquired with ISO scan imager (see Fig.4). The method is able to process a  $512 \times 512$  image in 250 ms (Matlab) with a non optimized implementation.

**References:** [32], [34]

Collaborators: Vincent Paveau and Cyril Cauchois (Innopys company), Hoai-Nam Nguyen.

# 7.3. Correlation-based method for membrane diffusion estimation during exocytosis in TIRFM

Participants: Ancageorgiana Caranfil, Charles Kervrann.

The dynamics of the plasma membrane of the cell is not fully understood yet; one of the crucial aspects to clarify is the diffusion process during exocytosis. Several image acquisition modalities exist, including TIRFM (Total Internal Reflection Fluorescence Microscopy), that have successfully been used to determine the successive steps of exocytosis. However, computing characteristic values for plasma membrane dynamics is problematic, as the experimental conditions have a strong influence on the obtained data, and a general model of molecular interaction dynamics cannot be determined.

In the PhD thesis of A. Caranfil, we have developed a computational approach to adapt the popular temporal image correlation spectroscopy (TICS) method to the analysis of a single fusing vesicle. The biophysical diffusion model parameters (for TfR protein) are estimated by an Approximate Bayesian Computing procedure which supplies the conditional expectation and maximum a posteriori estimators from temporal correlation data. Unlike TICS, our approach is robust to noise, estimation window size, spot location and non-uniform background. It can serve in biological studies investigating diffusion processes involved in exocytosis mechanisms.

Collaborators: Francois Waharte (UMR 144 CNRS-Institut Curie, PICT-IBiSA).



Figure 4. Three-color fluorescence image of eight tissue microarray cores. A region of interest of  $4.7 \times 2.8 mm^2$  was scanned using the fluorescence scanner named InnoScan 1100AL equipped with three excitation wavelengths (488nm, 532nm and 635nm) at the spatial resolution  $0.5\mu m^2$ / pixel, corresponding to an image of  $9544 \times 4704$  pixels. Two areas which are bordered by two blue and yellow boxes are selected for visual comparison. First row: full size image. Second and third rows: zoom-in views of two selected areas; from left to right: 3 synchronized colors (red (488nm), green (532nm) and blue (635nm) channels) displayed separately (courtesy of Innopsys).
#### 7.4. Classification of diffusion dynamics from particle trajectories

Participants: Vincent Briane, Charles Kervrann.

In this study, we are currently interested in describing the dynamics of particles inside live cell. Inference on the modes of mobility of molecules is central in cell biology since it reflects the interactions between the structures of the cell. In this work, we assume that the motions of particles follow a certain class of random process: the diffusion processes. Diffusions are stochastic processes with continuous paths and can model a large range of intracellular movements. Biophysicists distinguish three main types of diffusions, namely Brownian motion, superdiffusion and subdiffusion. These different diffusion processes correspond to distinct biological scenarios. A particle evolving freely inside the cytosol or along the plasma membrane is modelled by Brownian motion; the particle does not travel along any particular direction and can take a very long time to go to a precise area in the cell. Active intracellular transport can overcome this difficulty so that motion is faster in a given direction. In this case, particles are carried by molecular motors along microtubular filament networks and their motion is modelled with superdiffusion. Subdiffusion can be observed in two cases i/ when the particle is confined in a microdomain, ii/ when the particle is hindered by molecular crowding and encounters dynamic or fixed obstacles.

To address several issues in dynamics classification, we have developed a statistical test for classifying the observed trajectories into the three groups of diffusion of interest, namely Brownian motion, super-diffusion and subdiffusion. We have also designed an algorithm to detect the changes of dynamics along a single trajectory (see Fig. 5). We define the change points as the instants at which the particle switches from one diffusion type (Brownian motion, superdiffusion or subdiffusion) to another one. Finally, we have combined a clustering algorithm with our test procedure to identify micro domains, that is, zones where the particles are confined. Molecular interactions of great importance for the functioning of the cell take place in such areas.

Collaborators: Myriam Vimond (ENSAI Rennes),

Jean Salamero (UMR 144 CNRS-Institut Curie).



Figure 5. Change point detection on trajectories depicting neuronal mRNPs. The blue parts correspond to Brownian portions of the trajectory, red part to superdiffusive portions, green part to the subdiffusive portion.

# 7.5. Spatial statistics, point patterns, and colocalization in fluorescence imaging

Participants: Frédéric Lavancier, Charles Kervrann.



Figure 6. DSTORM acquisition of cells from hippocampi of mice expressing BDNF proteins (green channel) and vGlut (purple channel), with three zoomed-in regions (bottom). The colocalization regions identified by GcoPS are represented as white circles. The red rectangle represents the window used to find the colocalization hit shown as a red circle. The scale bars correspond to  $1\mu m$ 

In the context of bioimaging, colocalization refers to the detection of emissions from two or more fluorescent molecules within the same pixel of the image. This approach enables to quantify the protein-protein interactions inside the cell, just at the resolution limit of the microscope. It refers to the detection of emissions from two or more fluorescent molecules within the same pixel of the image. Colocalization is an open problem for which no satisfying solution has been found up to now. Accordingly, we proposed an objective, robust-to-noise colocalization method (GcoPS – Geo-coPositioning System)) which only requires the adjustment of a p-value that guarantees more reproducibility and more objective interpretation. It is based on the statistical analysis of the intersection (area/volume) between the two 2D or 3D binary segmented images. GcoPS handles 2D and 3D images, variable signal-to-noise ratios and any fluorescence image pair acquired with conventional or super-resolution microscopy (see Fig. 6). To our knowledge, no existing method offers the same robustness and precision level with such an easy control of the algorithm. In a recent study (internships 2017), we started to adapt this framework to analyze the spatiotemporal molecular interactions from set of 3D computed trajectories or motion vector fields (e.g., co-alignment), and then to fully quantify specific molecular machineries.

More generally, analysis of molecule and protein localization, of interactions and spatial distributions in living cells is helpful to understand functions in the cell and to compare spatialized phenotypes. This is also true with the emergence of single-molecule localization microscopy techniques (e.g., PALM), relying on the cumulative spatial localization of fluorescently tagged markers, and whose outputs are sets of spatial coordinates of single molecules. Accordingly, we were interested in the spatial distribution of single molecules that exhibit some randomness, regularity and spatial clustering (or aggregation) at large scales, while having a minimal distance between them. In that context, we theoretically studied several point processes able to represent the spatial organization of points. We focused on determinantal point processes (DDP), since they are able to describe spatial point patterns where nearby points repel or repulse each other. We also partly solved a 30 years old conjecture by proving the consistency of the likelihood procedure for a large class of Gibbs models (e.g., Strauss model, area-interaction model) which are commonly used models in practice. We extended the pseudo-

likelihood procedure to infinite range Gibbs interactions, and we proved its consistency and its asymptotic normality. All these models are now well understood and will be used in future works to analyse point patterns in cell imaging, generally described by Poisson point processes.

**References:** [30], [31], [35]

Collaborators: Jean Salamero and Liu Zengzhen (UMR 144 CNRS-Institut Curie),

David Dereudre (Laboratoire Paul Painlevé (UMR 8524), University of Lille 1), Jean-François Coeurjolly (Laboratoire Jean Kutzmann, University of Grenoble).

#### 7.6. Data assimilation and modeling of cell division mechanism

Participants: Ancageorgiana Caranfil, Charles Kervrann.

Nowadays, medical challenges demand a profound understanding of cellular mechanisms. Research in biology, biophysics and medical domain unravelled a significant part of the general processes occurring at the cellular level. It has enabled the understanding of much smaller scale processes, but our knowledge on these mechanisms is still limited as new, more complex issues need to be solved. In this context, we aim at understanding the role and interaction of the molecular key players at different scales, and their individual and collective impact on the global mechanism at the cell level. To this purpose, we have focused on the dynamics of the spindle during cell division mechanism. Our approach consists in creating a biophysical model for this mechanism, and uses data assimilation to adjust the model and optimally integrate the information from the observations. The overall spindle behaviour is led by the spindle poles behaviour. This year, we have proposed a new biophysical model for the posterior spindle pole functioning during metaphase and anaphase, that explains the oscillatory behaviour with a minimum number of parameters. Estimating the model parameters is ongoing, and will provide insights on molecular players role as well as guidance for future experiments to further investigate the dynamics of the spindle during cell division. First, we have focused on the temporal aspect. Spatial information on microtubules and molecular motors will be included in the model in the next part of this work.

Collaborators: Yann Le Cunff and Jacques Pécréaux (IGDR Institute of Genetics & Development of Rennes).

#### 7.7. Quantifying the spatial distribution of intracellular events

#### Participant: Charles Kervrann.

Automated processing of fluorescence microscopy data allows to quantify cell phenotypes in an objective and reproducible way. However, most computational methods are based on the complex combination of heterogeneous features expressing geometrical, morphological and frequency properties, which makes difficult to draw definitive biological conclusions. Additionally, most experimental designs pool together data coming from replicated experiments of a given condition, neglecting the biological variability between individual cells. Hence, we developed a generic and nonparametric density framework (QuantEv) to discriminate spatiotemporal distributions (using circular Earth mover's distance) of moving proteins detected by any appropriate algorithm. The main advantage of QuantEv is to robustly process 2D and 3D data, and accurately analyse homogeneous and heterogeneous populations. As proof-of-principle, we first quantitatively characterized protein trafficking of Rab6 positive membranes between the Golgi apparatus and the plasma membrane. Next, we demonstrated that Rab11 positive membranes uniformly distribute around the Endosomal Recycling Compartment (ERC), regardless of the cell shape. Finally, we showed that actin organization is cell shape dependent, and evaluated its influence on the distribution of exocytosis/recycling vesicles. QuantEv is a flexible method which enables to quantify any intracellular trafficking in 3D flat or rounded, constrained or non-constrained, adherent or non-adherent cells.

#### References: [36]

Collaborators: Thierry Pécot (Hollings Cancer Center, Medical Univ. South Carolina, Charleston, USA), Jean Salamero, Jérôme Boulanger and Liu Zengzhen (UMR 144 CNRS-Institut Curie).



Figure 7. Illustration of the cell division mechanism observed in fluorescence microscopy (A). Sketch of one centrosome and connected to microtubules in the cell (B), experiments and tracking of the two centrosomes (C and D), and simulation of centrosome oscillations (E).



Figure 8. Overview of QuantEv approach.

#### 7.8. 3D registration for correlative light-electron microscopy

Participants: Bertha Mayela Toledo Acosta, Patrick Bouthemy.

In recent years, correlative light and electron microscopy (CLEM) has become an attractive tool in the bioimaging field. Biologists can collect complementary information from light microscopy (LM) and electron microscopy (EM), respectively on the dynamics and on the structure of the cell. An overlay of the LM and EM images is needed to combine information from the LM and EM sources. We are developing a 3D automated CLEM method to register EM and LM image stacks. Given the significant gap between the field of view, position and orientation of the EM and LM stacks, it is not possible to estimate directly the 3D registration. We first compute a 2D maximum intensity projection (MPI) of the LM stack along the Z-axis, and we match 2D EM regions of interest (ROI), extracted from different EM slices, into the 2D LM-MPI image. From the resulting location candidates, we estimate with a robust criterion the 2D XY-shift to pre-align the LM and EM stacks. Afterwards, a 3D affine transformation between 3D-LM-ROI and 3D-EM–ROI can be estimated using mutual information. We successfully tested this framework on two first 3D correlative microscopy datasets.

Collaborators: Xavier Heiligenstein (UMR 144 CNRS-Institut Curie),

Grégoire Malandain (Inria, Morpheme EPC, Sophia-Antipolis).

#### 7.9. Fast optical flow methods for 3D fluorescence microscopy

Participants: Sandeep Manandhar, Patrick Bouthemy, Charles Kervrann.

Estimating motion of cells and of subcellular particles is crucial for deciphering cell mechanisms and understanding cell behaviors. Modern 3D light microscopy (LM) for cell biology enables to observe cell dynamics at a good resolution, in both space and time, motivating the development of 3D optical flow methods. However, the acquired 3D LM image sequences exhibit several specificities making 3D motion computation a difficult problem. We have defined an original and efficient two-stage estimation method for light microscopy image volumes. The method, developed in the frame of S. Manandhar PhD thesis, takes a pair of LM image volumes as input, segments the 2D slices of the source volume in super-pixels, and first estimates the 3D displacement vectors of the super-pixel centers. To this end, we have extended the well-known PatchMatch method to 3D volumes, where correspondences act between voxels. Both the propagation and the random search steps were adapted to 3D volumes. Then, a weighted interpolation has been designed to recover the dense 3D flow field for all the voxels, from the sparse 3D displacement field. The super-pixel segmentation is exploited to define the neighborhood for interpolation, and the interpolation weights take into account intensity edges and local motion differences to preserve flow discontinuities. The experimental results show good gain in execution speed, and accuracy evaluated in computer-generated 3D data with ground-truth. The results were promising on two real 3D LM image sequences supplied by USTW. The sequences depict blebbing in a MV3 cell (see Fig. 9). The cell membrane protrudes increasing the surface area of the cell. These protrusions, referred to as blebs, appear and disappear in interval of minutes, the bleb appearance corresponding to the stretching of a local region of the cell membrane. The total computation time was for the first sequence 163 seconds (resp. 101s for the second sequence), with 19 (resp. 49), 120 (resp. 44) and 24 (resp. 8) seconds for super-pixel generation, 3D patch matching, and interpolation respectively, on a computer with 2.8 GHz Intel i7 processor and 16 GB of RAM.

Collaborators: Philippe Roudot and Gaudenz Danuser (UTSW, Dallas, USA).

# 7.10. 3D Convolutional Neural Networks for macromolecule localization in cryo-electron tomograms of intact cells

Participants: Emmanuel Moebel, Charles Kervrann.

In this study, we focus on macromolecule localization and classification in cryo-electron tomography (CET) images. Biologists are in need for efficient methods to localize macro-molecules (e.g. ribosomes) in frozen cell samples. The high amount of noise and imaging artifacts are the reasons why very few computational methods exist for this task. In fact, the most used method today is template matching (TM) whose resulting



Figure 9. Illustration of 3D optical flow computation to analyze bleb deformation during cell migration in Bessel beam light sheet microscopy (input images by courtesy of Danuser lab, UTSW Dallas, USA).

score map comprises a high amount of false positives. Therefore, it is necessary to apply post-processing techniques (ROI selection, classification) in order to refine the localization results. We propose an alternative localization method to TM, based on a convolutional neural network (CNN). The idea is to propose a robust and more straight-forward approach, allowing to bypass the conventional processing chain. By using python toolboxes optimized for GPU computing (elektronn, keras), we are able to reach computation time much lower than the current approach. Results on synthetic data demonstrate the superiority of our approach compared to TM. In addition, we applied our method on experimental data in order to localize sub-classes of ribosomes (membrane-bound and cytoplasmic ribosomes), a task difficult to achieve with TM alone. We are currently in the process of publishing these results. Future perspectives include localizing smaller macro-molecules, like proteasomes.

**Collaborators:** Damien Larivière (Fondation Fourmentin-Guilbert), Julio Ortiz, Antonio Martinez (Max-Planck Institute, Martinsried, Germany).

#### 7.11. Estimation of parametric motion models with deep neural networks

Participants: Juan Manuel Perez Rua, Patrick Bouthemy.

We have proposed an end-to-end learning architecture for estimating a parametric motion model for a moving scene. We handle motion outliers by using the supervised training trick that is used by stacked denoising autoencoders. Here, we define motion outliers as regions of the image whose motion does not correspond with the estimated parametric motion model. In other words, we seek to find a parametrized dominant motion of the dynamic scene. We leverage stacked hourglass networks with a final hard-coded block corresponding to the global parametric motion model estimator. This block replaces the decoder part of a convolutional autoencoder network, and it is end-to-end trainable since it involves linear operations only. Moreover, the hard-wired decoder allows the network to output the values of the parametric motion model given an input moving scene, even when the supervision acts on optical flow maps and not the motion model values. This means that our network is able to provide, as a by-product, a concise code that can be used as motion descriptor.



Figure 10. Illustration of 3D CNN to localize ribosomes isolated in the cytoplasm and close to the cell membrane in cryo-electron tomography (courtesy of Max-Planck Institute, Martinsried, Germany).

Collaborators: Tomas Crivelli and Patrick Pérez (Technicolor).

#### 7.12. Motion saliency in video sequences

Participants: Léo Maczyta, Patrick Bouthemy.

Dynamic (or motion) saliency is a means to detect unexpected or rare dynamic behaviors in video sequences acquired by a stationary or a mobile imaging device. Finding salient dynamic information in each image of a sequence is indeed crucial in many situations. We aim to extract saliency only from motion information, and to exhibit salient motion in contrast to its space-time context with no prior on the nature of both. So far, we have investigated a simpler problem than saliency map estimation. We deal with the classification of each image of a sequence as dynamically salient or not, that is, containing salient motion or not. We have explored convolutional neural network (CNN). We have designed two different networks. The first one relies on two intensity images, the first input image and the second image warped with the parametric dominant motion estimated between the two input images. The second one takes as input the difference between the computed optical flow and parametric dominant flow.

Collaborators: Olivier Lemeur (EPC Sirocco, Inria Rennes - Bretagne Atlantique).

# 8. Bilateral Contracts and Grants with Industry

#### 8.1. Bilateral Contracts with Industry

#### 8.1.1. Contract with Innopsys: Tissue microarrays (TMA) image analysis

Participants: Charles Kervrann.

Collaborators: Vincent Paveau and Cyril Cauchois (Innopys company).

A three-year contract has been established with Innopsys in 2013 to support the PhD thesis of Hoai-Nam Nguyen. The objective was to investigate and develop methods and algorithms dedicated to fluorescence images acquired by scanners and devices designed by the company. In this project, we focused on localization and segmentation of fluorescence tissue microarrays (TMA) cores in very large 2D images, de-arraying of digital images and correction of grid deformation adapted to devices, correction of scanning artifacts to improve image reconstruction and deconvolution of fluorescence TMA images. The algorithms are currently embedded into software and hardware products designed by Innopsys.

# 8.1.2. Contract (CIFRE) with Technicolor: Semantically meaningful motion descriptors for video understanding

Participants: Juan Manuel Perez Rua, Patrick Bouthemy.

Collaborators: Tomas Crivelli and Patrick Pérez (Technicolor).

A three-year contract has been established with Technicolor in January 2015 for a CIFRE grant supporting the PhD thesis of Juan Manuel Pérez Rúa. The purpose was to investigate new methods for extracting meaningful mid-level motion-related descriptors that may help for the semantic discovery of the content. First, we addressed the occlusion detection problem and proposed a novel approach where occlusion is formulated in terms of visual reconstruction. Contrary to the usual approaches, the proposed alternative does not critically depend on a pre-computed, dense displacement field, while being shown to be more effective. Second, we developed two hierarchical motion segmentation methods involving a compositional motion representation. The first one follows a frame-based labeling approach which amounts to minimizing a global energy function. The second one is trajectory-based and relies on tree-structured learning and sparse coding.

#### 8.1.3. Contract with OBSYS: microscope set-up control and inverse problems in microscopy

Participants: Giovanni Petrazzuoli, Charles Kervrann.

Collaborators: Charles Gudeudry (OBSYS).

A two-year contract was established with OBSYS in 2016 for hiring an expert-engineer (12 months). The objective is to investigate and develop software for the control of a microscope set-up and the analysis of fluorescence images. Fast and robust algorithms have been especially developed to improve image reconstruction of 3D-TIRF microscope images. The algorithms will be embedded into platforms and devices designed by OBSYS. Giovanni Petrazzuoli has been hired in August 2017 on a full-time R&D engineer position in OBSYS (CDI). The collaboration with Inria will be pursued in 2018.

#### 8.2. Bilateral grants with industry

#### 8.2.1. Fourmentin-Guilbert Foundation: Macromolecule detection in cryo-electron tomograms

Participants: Emmanuel Moebel, Charles Kervrann.

Collaborator: Damien Larivière (Fourmentin-Guilbert Foundation).

A three-year contract was established with Fondation Fourmentin-Guilbert to partly support the PhD thesis of Emmanuel Moebel. The Fondation Fourmentin-Guilbert strives for building a virtual E. coli bacteria. Information about the position of macromolecules within the cell is necessary to achieve such a 3D molecularly-detailed model. The Fondation Fourmentin-Guilbert supports cutting-edge *in-situ* cryo-electron tomography combined with image processing at the Max-Planck Institute of Biochemistry to map the spatial distribution of the ribosomes, and obtain structural information on the complexes they form *in-situ* with cofactors and other ribosomes. The objective of the project is to explore novel methods from the field of 3D shape retrieval for identifying and counting macromolecules within a tomogram. This project is also supported by Région Bretagne.

# 9. Partnerships and Cooperations

#### 9.1. Regional Initiatives

**ENSAI-CREST:** Statistical methods and models for image registration, PhD thesis of Vincent Briane is cofunded by Inria and ENSAI-CREST and co-supervised by Myriam Vimond (ENSAI-CREST).

**Région Bretagne:** Identification, localization and enumeration of ribosomes within a tomogram by combining state-of-the-art denoising methods and object descriptor-based recognition (CATLAS, see Section 8.2.1) (PhD thesis of Emmnuel Moebel); motion saliency in video sequences (PhD thesis of Léo Maczyta).

**BioGenOuest:** Collaboration with S. Prigent (engineer) in charge of the organization of image processing services for Biogenouest bio-imaging facilities.

IGDR: Collaboration with J. Pecreaux, Y. Le Cunff (co-supervision of PhD thesis of A. Caranfil).

#### 9.2. National Initiatives

#### 9.2.1. France-BioImaging project

Participants: Charles Kervrann, Patrick Bouthemy.

The goal of the France-BioImaging project (http://france-bioimaging.org/) is to build a distributed coordinated French infrastructure for photonic and electronic cellular bioimaging, dedicated to innovation, training and technology transfer. High-computing capacities are needed to exhaustively analyse image flows. Serpico is cohead of the IPDM (Image Processing and Data Management) node of the FBI network composed of 6 nodes. In this context, we address the following scientific problems: i/ exhaustive analysis of bioimaging data sets; ii/ deciphering of key steps of biological mechanisms at organ, tissular, cellular and molecular levels through the systematic use of time-lapse 3D microscopy and image processing methods; iii/ storage and indexing of extracted and associated data and metadata through an intelligent data management system. Serpico recruited R&D engineers (2011-2016) to disseminate image processing software, to build the Mobyle@Serpico web portal and to manage the IGRIDA-Serpico cluster (200 nodes; batch scheduler: OAR; File management: Puppet/Git/Capistrano; OS: Linux Debian 7; User connexion: public ssh key) opened for end-users and dedicated to large scale computing and data sets processing (storage: 200 TeraBytes).

- Coordinator: CNRS (Jean Salamero, UMR 144 CNRS-Institut Curie).
- **Partners:** University of Paris-Diderot-Paris 7, Aix-Marseille University, University of Bordeaux, University of Montpellier, Institut Pasteur, Institut Curie, Inria, ENS Ulm, University of Paris Descartes, UPMC, Ecole Polytechnique, Inserm.
- **Funding:** Investissement d'Avenir Infrastructures Nationales en Biologie et Santé, ANR INBS-PIA 2011.
- Total amount:26 000 Keuros (Inria Serpico: 606 Keuros).

# 9.2.2. ANR DALLISH project (2016-2020): Data Assimilation and Lattice Light SHeet imaging for endocytosis/exocytosis pathway modeling in the whole cell

Participants: Charles Kervrann, Vincent Briane, Ancageorgiana Caranfil, Antoine Salomon.

Cutting-edge LLS microscopy represents the novel generation of 3D fluorescence microscopes dedicated to single cell analysis, generating extraordinarily high resolved and sharp, but huge 3D images and videos. One single live cell experiment in one single biological condition can result into up to one terabyte of data. The goal of the project is to develop new paradigms and computational strategies for image reconstruction and 3D molecule tracking/motion estimation. Furthermore, establishing correspondences between image-based measurements and features, stochastic motion models, and underlying biological and biophysical information remains a challenging task. In a larger perspective, the quantitative description of image data corresponding to protein transport will be a prerequisite for understanding the functioning of a cell in normal and pathological situations including cancer, viral infection and neurodegenerative diseases.

- Coordinator:Inria (Charles Kervrann)
- **Partners:**Inria (Serpico, Beagle, Fluminance teams), INRA MaIAGE Unit Jouy-en-Josas, Institut Curie (UMR 144 CNRS & U1143 Inserm UMR 3666) Paris
- Funding:ANR (Agence Nationale de la Recherche) PRC (Collaborative Research Project)
- Total amount:440 Keuros (Inria Serpico: 170 Keuros).

#### 9.3. European Initiatives

#### 9.3.1. Major European Organizations with which the Team have followed Collaborations

**ESFRI Euro-BioImaging initiative**: SERPICO participates in the ESFRI Euro-BioImaging project, one of the four new biomedical science projects in the roadmap of the European Strategic Forum on Research Infrastructures (ESFRI). The mission of Euro-BioImaging is to provide access, service and training to state-of-the-art imaging technologies and foster the cooperation and networking at the national and European level including multidisciplinary scientists, industry, regional, national and European authorities. SERPICO also participates to the French counterpart, the so-called "France-BioImaging" (FBI) network which gathers several outstanding cellular imaging centers (microscopy, spectroscopy, probe engineering and signal processing) as described in Section 9.2.1.

- Coordinator: EMBL (Jan Ellenberg, Heideldberg, Germany)
- Partners: 15 european countries in 2017
- Funding: Member states of the European Union

#### 9.4. International Initiatives

#### 9.4.1. Informal International Partners

Collaboration with Max-Planck Institute, Martinsried (Germany), Dr. Julio Ortiz and Antonio Martinez: Detection and segmentation of macromolecules in cryo-electron tomography (project in progress with Emmanuel Moebel and Charles Kervrann).

Collaboration with Aalborg University (Denmark), Prof. Rasmus Waagepetersen: Estimating equations for inhomogeneous determinantal point processes (project with Frédéric Lavancier).

#### 9.4.2. Inria Associate Teams Not Involved in an Inria International Labs

#### 9.4.2.1. CytoDI Inria Associated-Team

Title: Quantitative Imaging of Cytoskeleton Dynamics in 3D

International Partner:

University of Texas, SouthWestern Medical Center, Dallas (United States) - Gaudenz Danuser

Start year: 2016

See also: http://serpico.rennes.inria.fr/doku.php?id=research:cytodi

Participants: Sandeep Manandhar, Patrick Bouthemy, Charles Kervrann.

The main scientific goal of the Associated-Team is the spatiotemporal characterization and comparison of cytoskeleton networks involved in cell migration and observed through live cell imaging in three dimensions (3D). Those networks include the cytoskeleton, i.e., microtubules (MT), intermediate filaments (IF), dynamically resolvable by Bessel Beam Light Sheet fluorescent microscopy. The goal will be achieved through the design of local and global descriptors of the spatial conformation and deformation of the cytoskeleton. Subsequently, general metrics to compare and classify the MT and IF networks will be investigated. This study will be carried out on oncogenically transformed lung cancer epithelial cells.

The second meeting of the AT CytoDI took place in Rennes in July 2017 (visit of P. Roudot and K. Dean), to discuss and update current research direction and discuss scientific progress. Several meetings were organized with students (S. Manandhar, V. Briane, E. Moebel, T. Dubois, Q. Delannoy) to synchronize development in optical flow, co-orientation and visualization. The Danuser team focused on presenting recent imaging and analysis capacities as well as the current solution in development for the systematic analysis, contextualization and interpretation of 3D dynamics for quantitative biology.

#### 9.5. International Research Visitors

#### 9.5.1. Visits to International Teams

Emmanuel Moebel attended a summer school (one week): Signal Processing Meets Deep Learning (Capri, Italy, 4-8 september 2017).

Sandeep Manandhar attended a summer school (one week): VISion Understanding and Machine intelligence (Porto, Portugal, 7-14 July 2017).

## **10. Dissemination**

#### **10.1. Promoting Scientific Activities**

#### 10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

- Charles Kervrann is head of the organizing committee of the "Quantitative BioImaging" (QBI'2019) conference (Rennes, January 2019).

- Patrick Bouthemy was co-organizer of the BioImage Computing (BIC) workshop in conjunction with ICCV'2017 (Venice, Italy).

#### 10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

- Charles Kervrann: senior PC (Program Committee) of "BioImage Informatics 2017" conference, member of the scientific committee of the "Imaging the Cell 2017" conference (Rennes), member of the scientific committee of RFIAP'2018 (Marne-la-Vallée), Associated Editor for the ISBI'2018 conference, member of the scientific committee of the JIONC workshop (GdR Ondes, since 2014).

- Patrick Bouthemy: Associate Editor for the ISBI'2017 conference, member of the program committee of the IPTA'2017 conference.

#### 10.1.2.2. Reviewer

- Charles Kervrann: reviewer for ICIP'2017, ICASSP'2017, ISBI'2017, and SSVM'2017, PBIC'2017.

#### 10.1.3. Journal

#### 10.1.3.1. Member of the Editorial Boards

- Charles Kervrann is Associate Editor of the IEEE Signal Processing Letters journal.

- Patrick Bouthemy is co-editor in chief of the open access journal Frontiers in ICT, specialty Computer Image Analysis.

#### 10.1.3.2. Reviewer - Reviewing Activities

- Charles Kervrann: BMC Bioinformatics, Int. J. Computer Vision, SIAM J. on Imaging Sciences, Machine Vision and Applications.

#### 10.1.4. Invited Talks

- Charles Kervrann: Invited talk at II SLAS "HCS – High Contents Screening" conference (Madrid, Spain), "Connecting The Dots – Intelligent Trackers" workshop (LAL-Orsay), and GdR ImaBio (with V. Briane, University of Montpellier; with E. Moebel, Inria Rennes); Invited seminars at IDIAP (Martigny, Switzerland).

- Patrick Bouthemy: Invited speaker at 5ème colloque de la Société Française des Microscopies (Bordeaux).

- Frédéric Lavancier: Invited talks at GdR ImaBio-FBI "BioImage Informatics" (Rennes).

#### 10.1.5. Leadership within the Scientific Community

- Charles Kervrann is member of the IEEE BISP "Biomedical Image and Signal Processing" (https://signalprocessingsociety.org/get-involved/bio-imaging-and-signal-processing) committee (expert committee for ISBI, ICASSP and ICIP conferences). He is member of the executive board of the GdR MIV/ImaBio (2588 - Microscopie Fonctionnelle du Vivant) CNRS. He is member of the scientific committee of the Interdisciplinary MiFoBio School CNRS (http://gdr-miv.fr/mifobio2018/).

- Patrick Bouthemy is member of the board of AFRIF (Association Française pour la Reconnaissance et l'Interprétation des Formes).

#### 10.1.6. Scientific Expertise

- Charles Kervrann was expert for the international project evaluation in the framework of Luxembourg NRF (National Research Fund – RESCOM program). He was expert for the national project evaluation in the framework of DIM ELICIT Institut Pasteur - Région Ile-de-France in 2017. He was member of the selection committee for a Professor position at the University of Bordeaux (Section CNU 26, Mathematics) in 2017.

- Patrick Bouthemy was member of the evaluation committee of the ANR-DGA MALIN Challenge, member of the HCERES visiting committee of the DI-ENS (Département d'Informatique - Ecole Nationale supérieure Paris) (November 2017). He is member of the Research Committee of IMT Atlantique. He was member of the committee for professor promotion at IMT Atlantique in 2017.

#### 10.1.7. Research Administration

Charles Kervrann is member of the executive board of the project committee of the Inria Rennes
Bretagne Atlantique centre since 2010. He is Co-head of the "BioImage Informatics" node (ANR France-BioImaging project, National Infrastructure en Biologie et Santé) since 2011.

- Patrick Bouthemy has been head of Excellence Lab (Labex) CominLabs (http://www.cominlabs. ueb.eu) since April 2014. He is deputy member of the board of directors and member of the selection and validation committee of the Images & Réseaux competitivity cluster (http://images-et-reseaux. com/). He is deputy member of the board of directors of IRT (Technological Research Institute) b<>com (https://b-com.com/). He is the Inria representative in the steering committee of the DGA-Inria collaboration.

- Frédéric Lavancier is an elected member of CNU section 26.

#### **10.2. Teaching - Supervision - Juries**

#### 10.2.1. Teaching

Charles Kervrann:

- Engineer Degree: Genomics and Informatics, 4.5 hours, Ecole Nationale Supérieure des Mines de Paris.

- Master: From Bioimage Processing to BioImage Informatics, 5 hours, coordinator of the module (30 hours), Master 2 Research IRIV, Telecom-Physique Strasbourg and University of Strasbourg.

- Master: Geometric Modeling for Shapes and Images, 6 hours, Master 2 Research SISEA, University of Rennes 1.

- Engineer Degree and Master 2 Statistics and Mathematics: Statistical Models and Image Analysis, 37 hours + 15 hours (TP, Emmanuel Moebel), 3rd year, Ecole Nationale de la Statistique et de l'Analyse de l'Information (ENSAI), Rennes.

Patrick Bouthemy:

- Master: Analysis of Image Sequences, 18 hours, Master 2 Research SISEA, ISTIC & University of Rennes 1.

- Master: Video Indexing, 3 hours, Master 2 Research Computer Science, ISTIC & University of Rennes 1.

- Engineer Degree and Master 2 Research IRIV: Motion Analysis, 12 hours, Telecom-Physique Strasbourg & University of Strasbourg.

#### Frédéric Lavancier:

- Master: Linear Models, 36 hours, Master 2 Mathematics & Engineering, option Statistics, University of Nantes.

- Master: Time Series, 36 hours, Master 2 Mathematics & Engineering, option Statistics, University of Nantes.

#### 10.2.2. Supervision

- Arnaud Poinas, inference for inhomogeneous determinantal point processes, started in September 2016, supervised by Bernard Delyon and Frédéric Lavancier.

- Vincent Briane, statistical tests for particle trajectory analysis: application to intracellular imaging, started in October 2014, supervised by Charles Kervrann and Myriam Vimond (ENSAI-CREST).

- Bertha Mayela Toledo Acosta, methods and algorithms for 3D image registration and correlative microscopy, started in October 2014, supervised by Patrick Bouthemy and Charles Kervrann.

- Emmanuel Moebel, new strategies for the nonambiguous identification and enumeration of macromolecules in cryo-electron tomograms, started in November 2015, supervised by Charles Kervrann.

- Juan Manuel Perez Rua, hierarchical motion-based video analysis with applications to video postproduction, started in January 2015, supervised by Patrick Bouthemy in collaboration with Tomas Crivelli and Patrick Pérez (Technicolor).

- Ancageorgiana Caranfil, data assimilation methods for cell division mechanisms and molecule trafficking analysis, started in December 2016, supervised by Charles Kervrann and Yann Le Cunff.

- Sandeep Manandhar, optical flow methods for 3D fluorescence imaging, started in October 2016, supervised by Patrick Bouthemy and Charles Kervrann.

- Yunjiao Lu, intracellular dynamics and super-resolution imaging: analysis of bacteria wall at the molecular scale, started in October 2017, supervised by Charles Kervrann and Rut Carballido-Lopez.

- Antoine Salomon, statistical aggregation for image analysis in fluorescence microscopy and superresolution, started in November 2017, supervised by Charles Kervrann.

- Léo Maczyta, motion saliency in video sequences, started in October 2017, supervised by Patrick Bouthemy and Olivier Lemeur.

#### 10.2.3. Juries

*Referee of PhD thesis:* F. Zhou (University of Oxford, supervised by J. Rittscher) [P. Bouthemy], R. Giraud (University of Bordeaux, supervised by N. Papadakis) [P. Bouthemy], C. Jiang (University of Bourgogne, supervised by C. Demonceaux) [P. Bouthemy], A. Samacoïts (Institut Pasteur - UPMC, supervised by C. Zimmer) [C. Kervrann].

*President of PhD thesis jury:* F. Deslandes (INRA, supervised by B. Laroche), I. Ahmet (University of Rennes 1, supervised by T. Furon).

# 11. Bibliography

#### Major publications by the team in recent years

- [1] A. BASSET, J. BOULANGER, J. SALAMERO, P. BOUTHEMY, C. KERVRANN. Adaptive spot detection with optimal scale selection in fluorescence microscopy images, in "IEEE Transactions on Image Processing", November 2015, vol. 24, n<sup>o</sup> 11, 16 [DOI: 10.1109/TIP.2015.2450996], https://hal.inria.fr/hal-01248290.
- [2] A. BASSET, P. BOUTHEMY, J. BOULANGER, F. WAHARTE, J. SALAMERO, C. KERVRANN. An extended model of vesicle fusion at the plasma membrane to estimate protein lateral diffusion from TIRF microscopy images, in "BMC Bioinformatics", 2017, vol. 18, n<sup>o</sup> 1, 352 [DOI: 10.1186/s12859-017-1765-Y], https:// hal.inria.fr/hal-01561310.
- [3] J. BOULANGER, A. GIDON, C. KERVRANN, J. SALAMERO. *A patch-based method for repetitive and transient event detection in fluorescence imaging*, in "PLoS ONE", Oct 2010, vol. 5, n<sup>o</sup> 10 [DOI: 10.1371/JOURNAL.PONE.0013190].
- [4] J. BOULANGER, C. KERVRANN, P. BOUTHEMY. Space-time adaptation for patch based image sequence restoration, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2007, vol. 29, n<sup>o</sup> 6, p. 1096–1102.
- [5] J. BOULANGER, C. KERVRANN, P. BOUTHEMY.A simulation and estimation framework for intracellular dynamics and trafficking in video-microscopy and fluorescence imagery, in "Medical Image Analysis", 2009, vol. 13, p. 132–142.
- [6] J. BOULANGER, C. KERVRANN, P. BOUTHEMY, P. ELBAU, J.-B. SIBARITA, J. SALAMERO.Patch-based nonlocal functional for denoising fluorescence microscopy image sequences, in "IEEE Transactions on Medical Imaging", Feb 2010, vol. 29, n<sup>o</sup> 2, p. 442-453 [DOI: 10.1109/TMI.2009.2033991].
- [7] P.M. CARLTON, J. BOULANGER, C. KERVRANN, J.-B. SIBARITA, J. SALAMERO, S. GORDON-MESSER, J. HABER, S. HAASE, L. SHAO, L. WINOTO, A. MATSUDA, P. KNER, S. USAWA, Y. STRUKOV, M. GUSTAFSSON, Z. KAM, D. AGARD, J.W. SEDAT.*Fast live simultaneous multiwavelength fourdimensional optical microscopy*, in "Proc Natl Acad Sci USA", Sep 2010, vol. 107, n<sup>o</sup> 37, p. 16016-16022 [DOI: 10.1073/PNAS.1004037107].
- [8] N. CHENOUARD, I. SMAL, F. DE CHAUMONT, M. MASKA, I. F. SBALZARINI, Y. GON, J. CARDINALE, C. CARTHEL, S. CORALUPPI, M. WINTER, A. R. COHEN, W. J. GODINEZ, K. ROHR, Y. KALAIDZIDIS, L. LIANG, J. DUNCAN, H. SHEN, Y. XU, K. MAGNUSSON, J. JALDEN, H. M. BLAU, P. PAUL-GILLOTEAUX, P. ROUDOT, C. KERVRANN, F. WAHARTE, J.-Y. TINEVEZ, S. L. SHORTE, J. WILLEMSE, K. CELLER, G. P. VAN WEZEL, H.-W. DAN, Y.-S. TSAI, C. ORTIZ DE SOLORZANO, J.-C. OLIVO-MARIN, E. MEIJERING. *Objective comparison of particle tracking methods*, in "Nature Methods", 2014, vol. 11, n<sup>o</sup> 3, p. 281–289 [DOI: 10.1038/NMETH.2808], https://hal.inria.fr/hal-00932869.
- [9] T. CRIVELLI, B. CERNUSCHI-FRIAS, P. BOUTHEMY, J.-F. YAO. Motion Textures: Modeling, Classification, and Segmentation Using Mixed-State, in "SIAM Journal on Imaging Sciences", December 2013, vol. 6, n<sup>o</sup> 4, p. 2484-2520 [DOI: 10.1137/120872048], https://hal.inria.fr/hal-00931667.

- [10] A. GIDON, S. BARDIN, B. CINQUIN, J. BOULANGER, F. WAHARTE, L. HÉLIOT, H. DE LA SALLE, D. HANAU, C. KERVRANN, B. GOUD, J. SALAMERO. A Rab11A/myosin Vb/Rab11-FIP2 complex frames two late recycling steps of langerin from the ERC to the plasma membrane, in "Traffic", April 2012, vol. 13, n<sup>o</sup> 6, p. 815-833 [DOI: 10.1111/J.1600-0854.2012.01354.x], https://hal.inria.fr/hal-00782005.
- [11] C. KERVRANN, J. BOULANGER. *Optimal spatial adaptation for patch-based image denoising*, in "IEEE Transactions on Image Processing", 2006, vol. 15, n<sup>o</sup> 10, p. 2866–2878.
- [12] C. KERVRANN, J. BOULANGER. Local adaptivity to variable smoothness for exemplar-based image denoising and representation, in "International Journal of Computer Vision", August 2008, vol. 79, n<sup>o</sup> 1, p. 45–69.
- [13] C. KERVRANN. PEWA: Patch-based Exponentially Weighted Aggregation for image denoising, in "NIPS -Neural Information Processing Systems", Montreal, Canada, Neural Information Processing Systems Foundation, December 2014, https://hal.inria.fr/hal-01103358.
- [14] T. PÉCOT, P. BOUTHEMY, J. BOULANGER, A. CHESSEL, S. BARDIN, J. SALAMERO, C. KERVRANN. Background Fluorescence Estimation and Vesicle Segmentation in Live Cell Imaging with Conditional Random Fields, in "IEEE Transactions on Image Processing", February 2015, vol. 24, n<sup>o</sup> 2, 14 [DOI: 10.1109/TIP.2014.2380178], https://hal.inria.fr/hal-01103126.

#### **Publications of the year**

#### **Articles in International Peer-Reviewed Journal**

- [15] A. BASSET, P. BOUTHEMY, J. BOULANGER, F. WAHARTE, J. SALAMERO, C. KERVRANN. An extended model of vesicle fusion at the plasma membrane to estimate protein lateral diffusion from TIRF microscopy images, in "BMC Bioinformatics", 2017, vol. 18, n<sup>o</sup> 1, 352 [DOI: 10.1186/s12859-017-1765-Y], https:// hal.inria.fr/hal-01561310.
- [16] J.-F. COEURJOLLY, F. LAVANCIER. Parametric estimation of pairwise Gibbs point processes with infinite range interaction, in "Bernoulli", February 2017, vol. 23, n<sup>o</sup> 2, p. 1299-1334, https://arxiv.org/abs/1412.2557 [DOI: 10.3150/15-BEJ779], https://hal.archives-ouvertes.fr/hal-01092225.
- [17] M. JAIN, J. VAN GEMERT, H. JÉGOU, P. BOUTHEMY, C. SNOEK. *Tubelets: Unsupervised action propos*als from spatiotemporal super-voxels, in "International Journal of Computer Vision", June 2017, p. 1–25, https://arxiv.org/abs/1607.02003 [DOI: 10.1007/s11263-017-1023-9], https://hal.archives-ouvertes.fr/hal-01561056.
- [18] Q. JIN, I. GRAMA, C. KERVRANN, Q. LIU. Non-local means and optimal weights for noise removal, in "SIAM Journal on Imaging Sciences", December 2017, https://hal.inria.fr/hal-01575918.
- [19] J.-M. PÉREZ-RÚA, A. BASSET, P. BOUTHEMY. Detection and Localization of Anomalous Motion in Video Sequences from Local Histograms of Labeled Affine Flows, in "Frontiers in information and communication technologies", May 2017 [DOI: 10.3389/FICT.2017.00010], https://hal.archives-ouvertes.fr/hal-01561069.
- [20] P. ROUDOT, L. DING, K. JAQAMAN, C. KERVRANN, G. DANUSER. *Piecewise-stationary motion modeling and iterative smoothing to track heterogeneous particle motions in dense environments*, in "IEEE Transactions on Image Processing", June 2017, vol. 26, n<sup>o</sup> 11, p. 5395 5410 [DOI : 10.1109/TIP.2017.2707803], https://hal.inria.fr/hal-01575754.

#### **Invited Conferences**

[21] P. BOUTHEMY.An automatic image-based registration method for correlative light-electron microscopy, in "15ème colloque de la Société Française des Microscopies", Bordeaux, France, July 2017, https://hal.inria.fr/ hal-01561172.

#### **International Conferences with Proceedings**

- [22] D. FORTUN, N. DEBROUX, C. KERVRANN. Spatially-variant kernel for optical flow under low signal-tonoise ratios: application to microscopy, in "ICCV Workshop - BioImage Computing (BIC)", Venice, Italy, IEEE, October 2017, 9, https://hal.inria.fr/hal-01575720.
- [23] O. MIKSIK, J.-M. PÉREZ-RÚA, P. H. S. TORR, P. PEREZ.ROAM: a Rich Object Appearance Model with Application to Rotoscoping, in "2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)", Honolulu, United States, July 2017, https://hal.archives-ouvertes.fr/hal-01576017.
- [24] B. M. TOLEDO ACOSTA, A. BASSET, P. BOUTHEMY, C. KERVRANN. Multi-scale spot segmentation with selection of image scales, in "ICASSP 2017 - The 42nd IEEE International Conference on Acoustics, Speech and Signal Processing", New Orleans, United States, March 2017, 5 [DOI : 10.1109/ICASSP.2017.7952489], https://hal.inria.fr/hal-01561164.

#### **Conferences without Proceedings**

- [25] V. BRIANE, M. VIMOND, C. KERVRANN. Classification of diffusion dynamics from particle trajectories, in "QBI - Quantitative BioImaging", College-Station, United States, January 2017, https://hal.inria.fr/hal-01575467.
- [26] V. BRIANE, M. VIMOND, C. KERVRANN. Robust classification of particle tracks for characterization of diffusion and dynamics in fluorescence microscopy, in "FOM - Focus on Microscopy", Bordeaux, France, April 2017, https://hal.inria.fr/hal-01575465.

#### **Other Publications**

- [27] R. BARDENET, F. LAVANCIER, X. MARY, A. VASSEUR. On a few statistical applications of determinantal point processes, 2017, To appear in ESAIM: proceedings and surveys, https://hal.archives-ouvertes.fr/hal-01580353.
- [28] V. BRIANE, C. KERVRANN, M. VIMOND. *A Statistical Analysis of Particle Trajectories in Living Cells*, June 2017, working paper or preprint, https://hal.inria.fr/hal-01557705.
- [29] F. LAVANCIER, J.-F. COEURJOLLY. *Approximation intensity for pairwise interaction Gibbs point processes using determinantal point processes*, December 2017, working paper or preprint, https://hal.archives-ouvertes. fr/hal-01659150.
- [30] F. LAVANCIER, T. PÉCOT, L. ZENGZHEN, C. KERVRANN. A Fast Automatic Colocalization Method for 3D Live Cell and Super-Resolution Microscopy, August 2017, working paper or preprint, https://hal.inria.fr/hal-01577118.
- [31] F. LAVANCIER, P. ROCHET. A tutorial on estimator averaging in spatial point process models, March 2017, https://arxiv.org/abs/1607.00864 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01341125.

- [32] H.-N. NGUYEN, V. PAVEAU, C. CAUCHOIS, C. KERVRANN. A variational method for dejittering large fluorescence line scanner images, August 2017, working paper or preprint, https://hal.inria.fr/hal-01577139.
- [33] H.-N. NGUYEN, V. PAVEAU, C. CAUCHOIS, C. KERVRANN.*ATMAD : robust image analysis for Automatic Tissue MicroArray De-arraying*, August 2017, working paper or preprint, https://hal.inria.fr/hal-01577238.
- [34] H.-N. NGUYEN, V. PAVEAU, C. CAUCHOIS, C. KERVRANN.Generalized Sparse Variation Regularization for Large Fluorescence Image Deconvolution, October 2017, working paper or preprint, https://hal.inria.fr/ hal-01609810.
- [35] A. POINAS, B. DELYON, F. LAVANCIER. Mixing properties and central limit theorem for associated point processes, May 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01519096.
- [36] T. PÉCOT, L. ZENGZHEN, J. BOULANGER, J. SALAMERO, C. KERVRANN. QuantEv: quantifying the spatial distribution of intracellular events, August 2017, working paper or preprint, https://hal.inria.fr/hal-01575913.

#### **References in notes**

- [37] D. AXELROD. Total Internal Reflection Fluorescent Microscopy in cell biology, in "Traffic", 2004, vol. 2, p. 4658–4668.
- [38] E. BETZIG, G. PATTERSON, R. SOUGRAT, O. LINDWASSER, S. OLENYCH, J. BONIFACINO, M. DAVID-SON, J. LIPPINCOTT-SCHWARTZ, H. HESS. *Imaging intracelluar fluorescent proteins at nanometer resolution*, in "Science", 2006, vol. 313, p. 1642–1645.
- [39] A. BUADES, B. COLL, J.-M. MOREL. A review of image denoising algorithms, with a new one, in "SIAM J. Multiscale Modeling & Simulation", 2005, vol. 4, n<sup>o</sup> 2, p. 490–530.
- [40] A. CHESSEL, B. CINQUIN, S. BARDIN, J. SALAMERO, C. KERVRANN. Computational geometry-based scale-space and modal image decomposition: application to light video-microscopy imaging, in "Conf. on Scale Space and Variational Methods (SSVM'09)", Voss, Norway, June 2009, p. 770–781.
- [41] T. CRIVELLI, P. BOUTHEMY, B. CERNUSCHI-FRIAS, J.-F. YAO. Simultaneous motion detection and background reconstruction with a conditional mixed-state Markov random field, in "International Journal of Computer Vision", 2011, vol. 94, n<sup>o</sup> 3, p. 295–316.
- [42] M. GUSTAFSSON, L. SHAO, P.M. CARLTON, R. WANG, I. GOLUBOVSKAYA, W. CANDE, D. AGARD, J.W. SEDAT.3D resolution doubling in wide-filed fluorescence microscopy by structured illumination, in "Biophysical J.", 2008, vol. 94, p. 4957–4970.
- [43] S. OZERE, P. BOUTHEMY, F. SPINDLER, P. PAUL-GILLOTEAUX, C. KERVRANN. Robust parametric stabilization of moving cells with intensity correction in light microscopy image sequences, in "10th International Symposium on Biomedical Imaging (ISBI)", IEEE, 2013, p. 468–471.
- [44] M. PINOT, V. STEINER, B. DEHAPIOT, B.-K. YOO, F. CHESNEL, L. BLANCHOIN, C. KERVRANN, Z. GUEROUI. Confinement induces actin flow in a meiotic cytoplasm, in "Proceedings National Academy of Sciences USA", July 2012, vol. 109, n<sup>O</sup> 29, p. 11705–11710 [DOI : 10.1073/PNAS.1121583109], http://hal.inria.fr/inserm-00717415.

# **Project-Team SIROCCO**

# Analysis representation, compression and communication of visual data

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: CNRS Université Rennes 1

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Vision, perception and multimedia interpretation

### **Table of contents**

1.	Personnel			
2.	Overall Objectives			
	2.1. Intr	oduction	850	
	2.2. An	alysis and modeling for compact representation and efficient processing	850	
	2.3. Res	storation, super-resolution, editing	850	
	2.4. Rej	presentation and compression of visual data	851	
	2.5. Dis	tributed processing and robust communication	851	
3.	Research	Program	851	
	3.1. Inti	roduction	851	
	3.2. Par	ameter Estimation and Inference	852	
	3.3. Dat	a Dimensionality Reduction	852	
	3.4. Per	ceptual Modelling	852	
	3.5. Co	ding theory	853	
4.	Application	Application Domains		
	4.1. Ov	erview	854	
	4.2. Co	npression of emerging imaging modalities	854	
	4.3. Net	worked visual applications	854	
	4.4. Edi	ting, post-production and computational photography	855	
5.	Highlight	s of the Year	855	
6.	New Softw	vare and Platforms	855	
	6.1. Sac	cadicModel	855	
	6.2. Qu	antizationAE	856	
	6.3. LF-	Inpainting	856	
	6.4. LF-	HLRA	856	
	6.5. GB	R-MVimages	857	
	6.6. Pla	tforms	857	
	6.6.1.	Light field editor	857	
	6.6.2.	Acquisition of multi-view sequences for Free viewpoint Television	857	
_	6.6.3. Light fields datasets		857	
7.	New Resu	New Results		
	7.1. An	alysis and modeling for compact representation	858	
	7.1.1.	Visual attention	858	
	7.1.2.	Saliency-based navigation in omnidirectional image	859	
	7.1.3.	Context-aware Clustering and Assessment of Photo Collections	859	
	7.1.4.	Light fields view extraction from lensiet images	860	
	/.1.3.	Super-rays for efficient Light fields processing	861	
	7.2. Rej	presentation and compression of large volumes of visual data	801	
	7.2.1.	Cloud-based image and video compression	801	
	7.2.2.	Rate-distortion optimized tone curves for HDR video compression	802	
	7.2.3.	Sparse image representation and deep learning for compression	862	
	7.2.4.	Graph-based multi-view video representation	802 862	
	1.2.3. 7 2 4	Light fields dimensionality reduction and compression	003	
	7.2.0.	Light herds dimensionantly reduction and compression	003 964	
	7.5. Kel	Transformation of the Rate distribution for color transfor	004 861	
	7.3.1. 72.2	Light field inpainting and edit propagation	004 861	
	7.3.2. 722	Light fields super resolution	004 865	
	7.3.3.	tributed processing and robust communication	00J 866	
	7/1	Interactive Coding for Navigation in 3D scenes (ICON 2D)	000 866	
	/.+.1.	incractive County for mavigation in 5D sectors (ICON 5D)	000	

	7.4.2.	Correlation model selection for interactive video communication	867
	7.4.3.	Optimal selection of reference sensors for spatially correlated data storage	867
8.	<b>Bilateral</b>	Contracts and Grants with Industry	<b>867</b>
	8.1.1.	CIFRE contract with Envivio/ Ericsson on LDR compatible HDR video coding	867
	8.1.2.	CIFRE contract with Harmonic on image analysis for HDR video compression	867
	8.1.3.	CIFRE contract with Technicolor on image collection analysis	868
	8.1.4.	CIFRE contract with Technicolor on light fields editing	868
	8.1.5.	CIFRE contract with Technicolor on light fields compressed representation	868
	8.1.6.	CIFRE contract with Technicolor on cloud-based image compression	868
9.	Partnersh	ips and Cooperations	<b>869</b>
	9.1. Reg	ional Initiatives	869
	9.2. Eur	opean Initiatives	869
	9.3. Inte	rnational Initiatives	869
	9.3.1.	Inria Associate Teams Not Involved in an Inria International Labs	869
	9.3.2.	Inria International Partners	870
10.	Dissemination		<b>870</b>
	10.1. Promoting Scientific Activities		870
	10.1.1.	Scientific Events Organisation	870
	10.1.2.	Journal	871
	10.1.3.	Invited Talks	871
	10.1.4.	Leadership within the Scientific Community	871
	10.1.5.	Scientific Expertise	871
	10.1.6.	Research Administration	872
	10.2. Tea	ching - Supervision - Juries	872
	10.2.1.	Teaching	872
	10.2.2.	Juries	872
11.	Bibliogra	ıphy	873

#### **Project-Team SIROCCO**

Creation of the Project-Team: 2012 January 01

#### **Keywords:**

#### **Computer Science and Digital Science:**

A5. - Interaction, multimedia and robotics

A5.3. - Image processing and analysis

A5.4. - Computer vision

A5.9. - Signal processing

#### **Other Research Topics and Application Domains:**

B6. - IT and telecom

## 1. Personnel

#### **Research Scientists**

Christine Guillemot [Team leader, Inria, Senior Researcher, HDR] Thomas Maugey [Inria, Researcher] Aline Roumy [Inria, Researcher]

#### **Faculty Member**

Olivier Le Meur [Univ de Rennes I, Associate Professor, until Sept. 2017, HDR]

#### **Technical Staff**

Laurent Guillo [CNRS, Research Engineer] Pierre Allain [Inria, from May 2017] Cédric Le Cam [Inria] Mikael Le Pendu [Inria, until Jul 2017, granted by Google Inc] Xin Su [Inria]

#### **PhD Students**

Jean Begaint [Technicolor] Pierre David [Inria] Elian Dib [Inria] Thierry Dumas [Inria] Simon Evain [Inria, from Dec 2017] David Gommelet [ERICSSON, until Dec. 2017] Fatma Hawary [Technicolor] Matthieu Hog [Technicolor] Hristina Hristova [Univ de Rennes 1, partly with FRVSense, until Sept. 2017] Dmitry Kuzovkin [Technicolor, partly with FRVSense, until Sept. 2017] Navid Mahmoudian Bidgoli [Inria] Mira Rizkallah [Univ de Rennes I] Jinglei Shi [Inria, from Oct 2017]

#### **Post-Doctoral Fellows**

Oriel Frigo [Inria, until Aug 2017] Xiaoran Jiang [Inria] Lara Younes [Inria, from Mar 2017]

#### Visiting Scientists

Ole Johannsen [University of Konstanz, from Nov. 2017 until Dec. 2017]

Chiara Galdi [Eurecom, May 2017] Ehsan Miandji [University of Linkoping, from June 2017 until Jul. 2017]

Administrative Assistant

Huguette Bechu [Inria]

# 2. Overall Objectives

#### 2.1. Introduction

The goal of the SIROCCO project-team is the design and development of algorithms and practical solutions in the areas of analysis, modelling, coding, and communication of visual data, i.e. natural 2D images, videos, multi-view sequences with a focus on recent imaging modalities such as high dynamic range imaging, light fields, and 360° videos. The objective is to cover several inter-dependent algorithmic problems of the end-toend transmission chain from the capturing, compression, transmission to the rendering of the visual data. The project-team activities are structured and organized around the following inter-dependent research axes:

- Analysis and modeling for compact representation and processing
- Representation and compression of visual data
- Restoration, super-resolution, editing
- Distributed processing and robust communication

While aiming at generic approaches, some of the solutions developed are applied to practical problems in partnership with industry (Technicolor, Ericsson, DGA) or in the framework of national projects. The application domains addressed by the project are networked visual applications taking into account their various requirements and needs in terms of compression, of resilience to channel noise and network adaptation, of advanced functionalities such as navigation, and of high quality rendering.

#### 2.2. Analysis and modeling for compact representation and efficient processing

Analysis and modeling of the visual data are crucial steps for a number of video processing problems: compression, loss concealment, denoising, inpainting, editing, content summarization and navigation. The focus is on the extraction of different cues such as scene geometry, edge, texture and motion, on the extraction of high-level features (GIST-like or epitomes), and on the study of computational models of visual attention, useful for different visual processing tasks. In relation to the above problems, the project-team considers various types of image modalities (natural 2D still and moving images, multi-view and multi-view plus depth video, high dynamic range images, light fields, 360° video content).

#### 2.3. Restoration, super-resolution, editing

Depending on the application and the type of content, various issues are being addressed, such as restoring the data to cope with limitations of sensors or to recover the signal from compression artefacts. The design of efficient methods for enhancing the data resolution in spatial, temporal or angular (in the case of light fields) dimensions is also part of the project-team objectives to cope with limitations of the visual sensors or for anti-aliased image-based rendering, e.g. from light fields. View synthesis remains a difficult ill-posed problem related to angular super-resolution of multi-view data and light fields which keeps evolving with in particular the use of deep learning techniques. Estimating the scene geometry and the scene flow for dynamic scenes is a critical step of the above processing problems. Editing the content, or restoring disoccluded areas in multi-view processing calls for efficient icnpainting techniques.

#### 2.4. Representation and compression of visual data

The objective is to develop algorithmic tools for constructing low-dimensional representations of various imaging modalities (2D images and videos, multi-view, light fields, ...). Our approach goes from the design of specific algorithmic tools to the development of complete compression algorithms. The algorithmic problems that we address include data dimensionality reduction, the design of compact representations using overcomplete dictionaries, transforms on graphs, or autoencoders based on deep learning architectures. Low rank and sparse models are the essence of transform coding and of many other processing methods (e.g., denoising, classification, registration, super-resolution, inpainting). Developing complete compression algorithms necessarily requires tackling topics beyond the issues of sparse data representation and dimensionality reduction. For example, the problem of spatial, inter-view or temporal prediction using deep learning techniques is also addressed. Finally, rate-distortion models for constructing rate-efficient representations with various features of scalability or low dynamic range compatibility in the case of high dynamic range content are also studied.

#### **2.5.** Distributed processing and robust communication

The goal is to develop theoretical and practical solutions for robust image and video transmission over heterogeneous and time-varying networks. The first objective is to construct coding tools that can adapt to heterogeneous networks. This includes the design of (i) sensing modules to measure network characteristics, of (ii) robust coding techniques and of (iii) error concealment methods for compensating for missing data at the decoder when erasures occur during the transmission. The first objective is thus to develop sensing and modeling methods which can recognize, model and predict the packets loss/delay end-to-end behaviour. Given the estimated and predicted network conditions (e.g. Packet Error Rate (PER)), the objective is then to adapt the data coding, protection and transmission scheme. However, the reliability of the estimated PER impacts the performance of FEC schemes. We investigate the problem of constructing codes which would be robust to channel uncertainty, i.e. which would perform well not only on a specific channel but also "universally", hence reducing the need for a feedback channel. This would be a significant advantage compared with rateless codes such as fountain codes which require a feedback channel. Another problem which we address is error concealment. This refers to the problem of estimating lost symbols from the received ones by exploiting spatial and/or temporal correlation within the video signal.

The availability of wireless camera sensors has also been spurring interest for a variety of applications ranging from scene interpretation, object tracking and security environment monitoring. In such camera sensor networks, communication energy and bandwidth are scarce resources, motivating the search for new distributed image processing and coding (Distributed Source Coding) solutions suitable for band and energy limited networking environments. In the past years, the team has developed a recognized expertise in the area of distributed source coding, which in theory allows for each sensor node to communicate losslessly at its conditional entropy rate without information exchange between the sensor nodes. However, distributed source coding (DSC) is still at the level of the proof of concept and many issues remain unresolved. The goal is thus to further address theoretical issues as the problem of modeling the correlation channel between sources, to further study the practicality of DSC in image coding and communication problems.

# 3. Research Program

#### 3.1. Introduction

The research activities on analysis, compression and communication of visual data mostly rely on tools and formalisms from the areas of statistical image modelling, of signal processing, of coding and information theory. However, the objective of better exploiting the Human Visual System (HVS) properties in the above goals also pertains to the areas of perceptual modelling and cognitive science. Some of the proposed research axes are also based on scientific foundations of computer vision (e.g. multi-view modelling and coding). We have limited this section to some tools which are central to the proposed research axes, but the design of complete compression and communication solutions obviously rely on a large number of other results in the areas of motion analysis, transform design, entropy code design, etc which cannot be all described here.

#### **3.2.** Parameter Estimation and Inference

Bayesian estimation, Expectation-Maximization, stochastic modelling

Parameter estimation is at the core of the processing tools studied and developed in the team. Applications range from the prediction of missing data or future data, to extracting some information about the data in order to perform efficient compression. More precisely, the data are assumed to be generated by a given stochastic data model, which is partially known. The set of possible models translates the a priori knowledge we have on the data and the best model has to be selected in this set. When the set of models or equivalently the set of probability laws is indexed by a parameter (scalar or vectorial), the model is said parametric and the model selection resorts to estimating the parameter. Estimation algorithms are therefore widely used at the encoder to analyze the data. In order to achieve high compression rates, the parameters are usually not sent and the decoder has to jointly select the model (i.e. estimate the model parameters) and extract the information of interest.

#### **3.3. Data Dimensionality Reduction**

Manifolds, locally linear embedding, non-negative matrix factorization, principal component analysis

A fundamental problem in many data processing tasks (compression, classification, indexing) is to find a suitable representation of the data. It often aims at reducing the dimensionality of the input data so that tractable processing methods can then be applied. Well-known methods for data dimensionality reduction include principal component analysis (PCA) and independent component analysis (ICA). The methodologies which will be central to several proposed research problems will instead be based on sparse representations, on locally linear embedding (LLE) and on the "non negative matrix factorization" (NMF) framework.

The objective of sparse representations is to find a sparse approximation of a given input data. In theory, given  $A \in \mathbb{R}^{m \times n}$ , m < n, and  $\mathbf{b} \in \mathbb{R}^m$  with  $m \ll n$  and A is of full rank, one seeks the solution of  $\min\{\|\mathbf{x}\|_0 : A\mathbf{x} = \mathbf{b}\}$ , where  $\|\mathbf{x}\|_0$  denotes the  $L_0$  norm of x, i.e. the number of non-zero components in z. There exist many solutions x to Ax = b. The problem is to find the sparsest, the one for which x has the fewest non zero components. In practice, one actually seeks an approximate and thus even sparser solution which satisfies  $\min\{\|\mathbf{x}\|_0 : \|A\mathbf{x} - \mathbf{b}\|_p \le \rho\}$ , for some  $\rho \ge 0$ , characterizing an admissible reconstruction error. The norm p is usually 2, but could be 1 or  $\infty$  as well. Except for the exhaustive combinatorial approach, there is no known method to find the exact solution under general conditions on the dictionary A. Searching for this sparsest representation is hence unfeasible and both problems are computationally intractable. Pursuit algorithms have been introduced as heuristic methods which aim at finding approximate solutions to the above problem with tractable complexity.

Non negative matrix factorization (NMF) is a non-negative approximate data representation <sup>0</sup>. NMF aims at finding an approximate factorization of a non-negative input data matrix V into non-negative matrices Wand H, where the columns of W can be seen as *basis vectors* and those of H as coefficients of the linear approximation of the input data. Unlike other linear representations like PCA and ICA, the non-negativity constraint makes the representation purely additive. Classical data representation methods like PCA or Vector Quantization (VQ) can be placed in an NMF framework, the differences arising from different constraints being placed on the W and H matrices. In VQ, each column of H is constrained to be unitary with only one non-zero coefficient which is equal to 1. In PCA, the columns of W are constrained to be orthonormal and the rows of H to be orthogonal to each other. These methods of data-dependent dimensionality reduction will be at the core of our visual data analysis and compression activities.

#### 3.4. Perceptual Modelling

Saliency, visual attention, cognition

<sup>&</sup>lt;sup>0</sup>D.D. Lee and H.S. Seung, "Algorithms for non-negative matrix factorization", Nature 401, 6755, (Oct. 1999), pp. 788-791.

The human visual system (HVS) is not able to process all visual information of our visual field at once. To cope with this problem, our visual system must filter out irrelevant information and reduce redundant information. This feature of our visual system is driven by a selective sensing and analysis process. For instance, it is well known that the greatest visual acuity is provided by the fovea (center of the retina). Beyond this area, the acuity drops down with the eccentricity. Another example concerns the light that impinges on our retina. Only the visible light spectrum lying between 380 nm (violet) and 760 nm (red) is processed. To conclude on the selective sensing, it is important to mention that our sensitivity depends on a number of factors such as the spatial frequency, the orientation or the depth. These properties are modeled by a sensitivity function such as the Contrast Sensitivity Function (CSF).

Our capacity of analysis is also related to our visual attention. Visual attention which is closely linked to eye movement (note that this attention is called *overt* while the covert attention does not involve eye movement) allows us to focus our biological resources on a particular area. It can be controlled by both top-down (i.e. goal-directed, intention) and bottom-up (stimulus-driven, data-dependent) sources of information <sup>0</sup>. This detection is also influenced by prior knowledge about the environment of the scene <sup>0</sup>. Implicit assumptions related to prior knowledge or beliefs play an important role in our perception (see the example concerning the assumption that light comes from above-left). Our perception results from the combination of prior beliefs with data we gather from the environment. A Bayesian framework is an elegant solution to model these interactions <sup>0</sup>. We define a vector  $\vec{v}_l$  of local measurements (contrast of color, orientation, etc.) and vector  $\vec{v}_c$  of global and contextual features (global features, prior locations, type of the scene, etc.). The salient locations *S* for a spatial position  $\vec{x}$  are then given by:

$$S(\overrightarrow{x}) = \frac{1}{p(\overrightarrow{v}_l | \overrightarrow{v}_c)} \times p(s, \overrightarrow{x} | \overrightarrow{v}_c)$$
(53)

The first term represents the bottom-up salience. It is based on a kind of contrast detection, following the assumption that rare image features are more salient than frequent ones. Most of existing computational models of visual attention rely on this term. However, different approaches exist to extract the local visual features as well as the global ones. The second term is the contextual priors. For instance, given a scene, it indicates which parts of the scene are likely the most salient.

#### **3.5. Coding theory**

OPTA limit (Optimum Performance Theoretically Attainable), Rate allocation, Rate-Distortion optimization, lossy coding, joint source-channel coding multiple description coding, channel modelization, oversampled frame expansions, error correcting codes.

Source coding and channel coding theory  $^{0}$  is central to our compression and communication activities, in particular to the design of entropy codes and of error correcting codes. Another field in coding theory which has emerged in the context of sensor networks is Distributed Source Coding (DSC). It refers to the compression of correlated signals captured by different sensors which do not communicate between themselves. All the signals captured are compressed independently and transmitted to a central base station which has the capability to decode them jointly. DSC finds its foundation in the seminal Slepian-Wolf  $^{0}$  (SW) and Wyner-Ziv  $^{0}$  (WZ) theorems. Let us consider two binary correlated sources X and Y. If the two coders communicate, it is well known from Shannon's theory that the minimum lossless rate for X and Y is given by the joint

<sup>&</sup>lt;sup>0</sup>L. Itti and C. Koch, "Computational Modelling of Visual Attention", Nature Reviews Neuroscience, Vol. 2, No. 3, pp. 194-203, 2001.

<sup>&</sup>lt;sup>0</sup>J. Henderson, "Regarding scenes", Directions in Psychological Science, vol. 16, pp. 219-222, 2007.

<sup>&</sup>lt;sup>0</sup>L. Zhang, M. Tong, T. Marks, H. Shan, H. and G.W. Cottrell, "SUN: a Bayesian framework for saliency using natural statistics", Journal of Vision, vol. 8, pp. 1-20, 2008.

<sup>&</sup>lt;sup>0</sup>T. M. Cover and J. A. Thomas, Elements of Information Theory, Second Edition, July 2006.

 <sup>&</sup>lt;sup>0</sup>D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources." IEEE Transactions on Information Theory, 19(4), pp. 471-480, July 1973.
<sup>0</sup>A. Wyner and J. Ziv, "The rate-distortion function for source coding ith side information at the decoder." IEEE Transactions on

<sup>&</sup>lt;sup>1</sup>A. Wyner and J. Ziv, "The rate-distortion function for source coding ith side information at the decoder." IEEE Transactions on Information Theory, pp. 1-10, January 1976.

entropy H(X, Y). Slepian and Wolf have established in 1973 that this lossless compression rate bound can be approached with a vanishing error probability for long sequences, even if the two sources are coded separately, provided that they are decoded jointly and that their correlation is known to both the encoder and the decoder.

In 1976, Wyner and Ziv considered the problem of coding of two correlated sources X and Y, with respect to a fidelity criterion. They have established the rate-distortion function  $R *_{X|Y}(D)$  for the case where the side information Y is perfectly known to the decoder only. For a given target distortion D,  $R *_{X|Y}(D)$  in general verifies  $R_{X|Y}(D) \leq R *_{X|Y}(D) \leq R_X(D)$ , where  $R_{X|Y}(D)$  is the rate required to encode X if Y is available to both the encoder and the decoder, and  $R_X$  is the minimal rate for encoding X without SI. These results give achievable rate bounds, however the design of codes and practical solutions for compression and communication applications remain a widely open issue.

# 4. Application Domains

#### 4.1. Overview

The application domains addressed by the project are:

- Compression with advanced functionalities of various imaging modalities
- Networked multimedia applications taking into account needs in terms of user and network adaptation (e.g., interactive streaming, resilience to channel noise)
- Content editing, post-production, and computational photography.

#### 4.2. Compression of emerging imaging modalities

Compression of visual content remains a widely-sought capability for a large number of applications. This is particularly true for mobile applications, as the need for wireless transmission capacity will significantly increase during the years to come. Hence, efficient compression tools are required to satisfy the trend towards mobile access to larger image resolutions and higher quality. A new impulse to research in video compression is also brought by the emergence of new formats beyond High Definition TV (HDTV) towards high dynamic range (higher bit depth, extended colorimetric space), or of formats for immersive displays allowing panoramic viewing, Free Viewpoint Video (FVV) and 3DTV.

Different video data formats and technologies are envisaged for interactive and immersive 3D video applications using omni-directional videos, stereoscopic or multi-view videos. The "omni-directional video" set-up refers to 360-degree view from one single viewpoint or spherical video. Stereoscopic video is composed of two-view videos, the right and left images of the scene which, when combined, can recreate the depth aspect of the scene. A multi-view video refers to multiple video sequences captured by multiple video cameras and possibly by depth cameras. Associated with a view synthesis method, a multi-view video allows the generation of virtual views of the scene from any viewpoint. This property can be used in a large diversity of applications, including Three-Dimensional TV (3DTV), and Free Viewpoint Video (FTV). In parallel, the advent of a variety of heterogeneous delivery infrastructures has given momentum to extensive work on optimizing the end-to-end delivery QoS (Quality of Service). This encompasses compression capability but also capability for adapting the compressed streams to varying network conditions. The scalability of the video content compressed representation and its robustness to transmission impairments are thus important features for seamless adaptation to varying network conditions and to terminal capabilities.

#### 4.3. Networked visual applications

The emergence of multi-view auto-stereoscopic displays has spurred a recent interest for broadcast or Internet delivery of 3D video to the home. Multiview video, with the help of depth information on the scene, allows scene rendering on immersive stereo or auto-stereoscopic displays for 3DTV applications. This application sector suffers from an accommodation-vergence conflict which arises with conventional 3D displays (with or without glasses). Since each eye receives a single view, the eyes tend to focus on the display screen (accommodation), whereas the brain perceives the depth of 3D images due to the different views seen by each eye (vergence).

On the other hand, Free-viewpoint television (FTV) is a system for watching videos in which the user can choose its viewpoint freely and change it at anytime. To allow this navigation, many views are proposed and the user can navigate from one to the other. The goal of FTV is to propose an immersive sensation without the disadvantage of Three-dimensional television (3DTV). With FTV, a look-around effect is produced without any visual fatigue since the displayed images remain 2D. However, technical characteristics of FTV are large databases, huge numbers of users, and requests of subsets of the data, while the subset can be randomly chosen by the viewer. This requires the design of coding algorithms allowing such a random access to the pre-encoded and stored data which would preserve the compression performance of predictive coding. This research also finds applications in the context of Internet of Things in which the problem arises of optimally selecting both the number and the position of reference sensors and of compressing the captured data to be shared among a high number of users.

Broadband fixed (ADSL, ADSL2+) and mobile access networks with different radio access technologies (RAT) (e.g. 3G/4G, GERAN, UTRAN, DVB-H), have enabled not only IPTV and Internet TV but also the emergence of mobile TV and mobile devices with internet capability. A major challenge for next internet TV or internet video remains to be able to deliver the increasing variety of media (including more and more bandwidth demanding media) with a sufficient end-to-end QoS (Quality of Service) and QoE (Quality of Experience).

#### 4.4. Editing, post-production and computational photography

Editing and post-production are critical aspects in the audio-visual production process. Increased ways of "consuming" visual content also highlight the need for content repurposing as well as for higher interaction and editing capabilities. Content repurposing encompasses format conversion (retargeting), content summarization, and content editing. This processing requires powerful methods for extracting condensed video representations as well as powerful inpainting techniques. By providing advanced models, advanced video processing and image analysis tools, more visual effects, with more realism become possible. Our activies around light field imaging also find applications in computational photography which refers to the capability of creating photographic functionalities beyond what is possible with traditional cameras and processing tools.

# 5. Highlights of the Year

#### 5.1. Highlights of the Year

#### 5.1.1. Awards

BEST PAPERS AWARDS : [27] MMSP 2017 - IEEE 19th International Workshop on Multimedia Signal Processing. P. DAVID, M. LE PENDU, C. GUILLEMOT.

## 6. New Software and Platforms

#### 6.1. SaccadicModel

Saccadic model of visual attention KEYWORDS: Visual saliency maps - Visual scanpath FUNCTIONAL DESCRIPTION: Saliency models compute a saliency map from an input image. Saliency maps are a 2D map encoding the ability of every location to attract our gaze. There exist many models in the literature and tremendous progresses have been made. However, they remain quite limited when applied to natural scene exploration. Indeed, the vast majority of these models ignore fundamental properties of our visual system. The most important one is that they overlook the sequential and time-varying aspects of overt attention. Saccadic models aim to predict the visual scanpath itself, i.e. the series of fixations and saccades an observer would perform to sample the visual environment. We propose a new and efficient method to simulate the visual scanpath. It provides scanpaths in close agreement with human behavior and the model can be tailored to simulate scanpaths in specific conditions and for various observer profiles.

- Author: Olivier Le Meur
- Contact: Olivier Le Meur

#### 6.2. QuantizationAE

**KEYWORDS:** Compression - Machine learning

FUNCTIONAL DESCRIPTION: This code learns an autoencoder to compress images. The learning is performed under a rate-distortion criterion, and jointly learns a transform (the autoencoder) and the quantization step for target rate points. The code is organized as follows. It first builds a set of luminance images (B1) for the autoencoder training, a set of luminance images (B2) to analyze how the auto-encoder training advances and a set of luminance images (B3) to evaluate the auto-encoders in terms of rate-distortion. It then trains several auto-encoders using a rate-distortion criterion on the set B1. The quantization can be either fixed or learned during this training stage. The set B2 enables to periodically compute indicators to detect overfitting. It finally compares the auto-encoders in terms of rate-distortion can be either fixed or variable during this test.

- Participants: Aline Roumy, Christine Guillemot and Thierry Dumas
- Contact: Aline Roumy

#### 6.3. LF-Inpainting

Light field inpainting based on a low rank model

KEYWORDS: Light fields - Low rank models - Inpainting

FUNCTIONAL DESCRIPTION: This code implements a method for propagating the inpainting of the central view of a light field to all the other views. To this end, it also implements a new matrix completion algorithm, better suited to the inpainting application than existing methods. A first option does not require any depth prior, unlike most existing light field inpainting algorithms. The code also implements an extended version to better handle the case where the area to inpaint contains depth discontinuities.

- Participants: Mikael Le Pendu and Christine Guillemot
- Contact: Christine Guillemot

#### 6.4. LF-HLRA

#### Light fields homography-based low rank approximation

KEYWORDS: Compression - Light fields - Low rank models - Dimensionality reduction

FUNCTIONAL DESCRIPTION: This code jointly searches for homographies to align the views of an input light field together with the components of its low rank approximation model. The code either uses a global homography per view or multiple homographies, one per region, the region being extracted using depth information.

- Participants: Xiaoran Jiang, Mikael Le Pendu and Christine Guillemot
- Contact: Christine Guillemot

#### 6.5. GBR-MVimages

Graph-based Representation for multi-view and light field images

KEYWORDS: Light fields - Multi-View reconstruction - Graph

FUNCTIONAL DESCRIPTION: Graph-Based Representation (GBR) describes color and geometry of multiview or light field image content using a graph. The graph vertices represent the color information, while the edges represent the geometry information, i.e. the disparity, by connecting corresponding pixels in neighboring images.

- Participants: Xin Su and Thomas Maugey
- Contact: Thomas Maugey

#### 6.6. Platforms

#### 6.6.1. Light field editor

Participants: Pierre Allain, Laurent Guillo, Christine Guillemot.

As part of the ERC Clim project, the EPI Sirocco is developing a light field editor, a tool analogous to traditional image editors such as the GNU image manipulation program Gimp or the raster graphic editor Photoshop but dedicated to light fields. As input data, this tool accepts for instance sparse light fields acquired with High Density Camera Arrays (HDCA) or denser light fields captured with microlens array (MLA). Two kinds of features are provided. Traditional features such as changing the angle of view, refocusing or depth map extraction are or will be soon supported. More advanced features are being integrated in our tool as libraries we have developed, such as segmenation or inpainting. For instance, a segmentation on a specific subaperture/view of light fields can be propagated to all subapertures/views. Thus, the so-segmented objects or zones can be colourized or even removed, the emptied zone being then inpainted. The tool and libraries are developed in C++ and the graphical user interface relies on Qt.

#### 6.6.2. Acquisition of multi-view sequences for Free viewpoint Television

Participants: Cédric Le Cam, Laurent Guillo, Thomas Maugey.

The scientific and industrial community is nowadays exploring new multimedia applications using 3D data (beyond stereoscopy). In particular, Free Viewpoint Television (FTV) has attracted much attention in the recent years. In those systems, user can choose in real time its view angle from which he wants to observe the scene. Despite the great interest for FTV, the lack of realistic and ambitious datasets penalizes the research effort. The acquisition of such sequences is very costly in terms of hardware and working effort, which explains why no multi-view videos suitable for FTV has been proposed yet.

In 2017, in the context of the project ADT ATeP (funded by Inriahub), such datasets have been acquired and some calibration tools have been developed. First 40 omnidirectional cameras and their associated equipments have been acquired by the team (thanks to Rennes Metropole funding). We have first focused on the calibration of this camera, *i.e.*, the development of the relationship between a 3D point and its projection in the omnidirectional image. In particular, we have shown that the unified spherical model fits the acquired omnidirectional cameras. Second, we have developed tools to calibrate the cameras in relation to each other. Finally, we have made a capture of 3 multiview sequences that are currently in preparation for a sharing with the community (Fig. 1). This work has been published in [41].

#### 6.6.3. Light fields datasets

Participants: Pierre Allain, Christine Guillemot, Laurent Guillo.

The EPI Sirocco makes extensive use of light field datasets with sparse or dense contents provided by the scientific community to run tests. However, it has also generated its own natural and synthetic contents.



Figure 1. The 40 omnidirectional cameras positioned for an indoor scene capture.

Natural content has been created with Lytro cameras (the original first generation Lytro and the Lytro Illum) and is already available to the community (https://www.irisa.fr/temics/demos/lightField/CLIM/DataSoftware. html). The team also owns a R8 Raytrix plenoptic cameras with which still and video contents have been captured. Applications taking advantage of the Raytrix API have been developed to extract views from the Raytrix lightfield. The number of views per frame is configurable and can be set for instance to 3x3 or 9x9 according to the desired sparsity.

Synthetic content has been generated from the Sintel film (https://durian.blender.org/download/), which is a short computer animated film by the Blender institute, part of the Blender Foundation. A specific Blender add-on is used to extract views from a frame. As previously, the number of views is configurable. Synthetic contents present the advantage to provide a ground truth useful to evaluate how efficient our algorithms are to compute, for instance, the depth maps.

# 7. New Results

#### 7.1. Analysis and modeling for compact representation

3D modelling, light-fields, 3D meshes, epitomes, image-based rendering, inpainting, view synthesis

#### 7.1.1. Visual attention

#### Participant: Olivier Le Meur.

Visual attention is the mechanism allowing to focus our visual processing resources on behaviorally relevant visual information. Two kinds of visual attention exist: one involves eye movements (overt orienting) whereas the other occurs without eye movements (covert orienting). Our research activities deal with the understanding and modeling of overt attention.

**Saccadic model:** Since 2015, we have worked on saccadic model, which predicts the visual scanpaths of an observer watching a scene displayed onscreen. In 2016, we proposed a first improvement consisting in using spatially-variant and context-dependent viewing biases. We showed that the joint distribution of saccade amplitudes and orientations is significantly dependent on the type of visual stimulus. In addition, the joint distribution turns out to be spatially variant within the scene frame. This model outperforms state-of-the-art saliency models, and provides scanpaths in close agreement with human behavior. In [19], [35], we went further by showing that saccadic models are a flexible framework that can be tailored to emulate observer's viewing tendencies. More specifically, we tailored the proposed model to simulate visual scanpaths of 5 age groups of observers (i.e. adults, 8-10 y.o., 6-8 y.o., 4-6 y.o. and 2 y.o.). The key point is that the joint distribution of saccade amplitude and orientation is a visual signature specific to each age group, and can be

used to generate age-dependent scanpaths. Our age-dependent saccadic model does not only output humanlike, age-specific visual scanpaths, but also significantly outperforms other state-of-the-art saliency models. We demonstrated that the computational modelling of visual attention, through the use of saccadic model, can be efficiently adapted to emulate the gaze behavior of a specific group of observers.

Effects on Comics by Clustering Gaze Data: Comics are a compelling communication medium conveying a visual storytelling. With a smart mixture of text or/and other visual information, artists tell a story by drawing the viewer attention on specific areas. With the digital comics revolution (e.g. mobile comic and webcomic), we are witnessed a resurgence of interest for this art form. This new form of comics allows not only to tackle a wider audience but also new consumption methods. An open question in this endeavor is identifying where in a comic panel the effects should be placed. We proposed a fast, semi-automatic technique to identify effects-worthy segments in a comic panel by utilizing gaze locations as a proxy for the importance of a region. We took advantage of the fact that comic artists influence viewer gaze towards narrative important regions. By capturing gaze locations from multiple viewers, we can identify important regions. The key contribution is to leverage a theoretical breakthrough in the computer networks community towards robust and meaningful clustering of gaze locations into semantic regions, without needing the user to specify the number of clusters. We have developed a method based on the concept of relative eigen quality that takes a scanned comic image and a set of gaze points and produces an image segmentation. A variety of effects such as defocus, recoloring, stereoscopy, and animations has been demonstrated. We also investigated the use of artificially generated gaze locations from saliency models in place of actual gaze locations.

**Perceptual metric for perceptual transfer:** Color transfer between input and target images has raised a lot of interest in the past decade. Color transfer aims at modifying the look of an original image considering the illumination and the color palette of a reference image. It can be employed for image and video enhancement by simulating the appearance of a given image or a video sequence. Different color transfer methods often result in different output images. The process of determining the most plausible output image is difficult and requires, due to the lack of an objective metric, time-consuming and costly subjective experiments. To overcome this problem, we proposed a perceptual model for evaluating results from color transfer methods [31]. From a subjective experiment, involving several color transfer methods, we build a regression model with random forests to describe the relationship between a set of features (e.g. objective quality, saliency, etc.) and the subjective scores. An analysis and a cross-validation showed that the predictions of the proposed quality metric are highly accurate.

#### 7.1.2. Saliency-based navigation in omnidirectional image

Participants: Olivier Le Meur, Thomas Maugey.

Omnidirectional images describe the color information at a given position from all directions. Affordable 360° cameras have recently been developed leading to an explosion of the 360 degrees data shared on the social networks. However, an omnidirectional image does not contain interesting content everywhere. Some part of the images are indeed more likely to be looked at by some users than others. Knowing these regions of interest might be useful for 360° image compression, streaming, retargeting or even editing. In the work published in [25], a new approach based on 2D image saliency is proposed both to model the user navigation within a 360° image, and to detect which parts of an omnidirectional content might draw users' attention. A double cube projection is first used to put the saliency estimation in the classical 2D image framework. Consecutively, the saliency map serves as a support for the navigation estimation algorithm.

#### 7.1.3. Context-aware Clustering and Assessment of Photo Collections

Participants: Dmitry Kuzovkin, Olivier Le Meur.

To ensure that all important moments of an event are represented and that challenging scenes are correctly captured, both amateur and professional photographers often opt for taking large quantities of photographs. As such, they are faced with the tedious task of organizing large collections and selecting the best images among similar variants. Automatic methods assisting with this task are based on independent assessment approaches, evaluating each image apart from other images in the collection. However, the overall quality of





Figure 2. Example of the saliency map (left) and the estimated navigation (right).

photo collections can largely vary due to user skills and other factors. We explore the possibility of contextaware image quality assessment, where the photo context is defined using a clustering approach, and statistics of both the extracted context and the entire photo collection are used to guide identification of low-quality photos. We demonstrate that the proposed method is able to adapt flexibly to the nature of processed albums and to facilitate the task of image selection in diverse scenarios.

#### 7.1.4. Light fields view extraction from lenslet images

Participants: Pierre David, Christine Guillemot, Mikael Le Pendu.

Practical systems have recently emerged for the capture of real light fields which go from cameras arrays to single cameras mounted on moving gantries and plenoptic cameras. While camera arrays capture the scene from different viewpoints, hence with a large baseline, plenoptic cameras use an array of micro-lenses placed in front of the photosensor to separate the light rays striking each microlens into a small image on the photosensors pixels, and this way capture dense angular information with a small baseline. Extracting views from the raw lenslet data captured by plenoptic cameras involves several processing steps: devignetting which, with white images, aims at compensating for the loss of illumination at the periphery of the micro-lenses, color demosaicing, alignment of the sensor data with the micro-lens array, and converting the hexagonal sampling grid into a rectangular sampling grid. These steps are quite critical as they have a strong impact on the quality of the extracted sub-aperture images (views).

We have addressed two important steps of the view extraction from lenslet data: color demosaicing and alignment of the micro-lens array on the photosensor. We have developed a new method guided by a white lenslet image for color demosaicing of raw lenslet data [27](best paper award). The white lenslet image gives measures of confidence on the color values which are then used to weight the color samples interpolation (see Fig.3. Similarly, the white image is used to guide the interpolation performed in the alignment of the micro-len arrays on the photosensor. The method significantly decreases the crosstalk artefacts from which suffer existing methods.



*Figure 3. (a) is the raw image we want to demosaic, (b) is a mask which holds every pixel belonging to the same lenslet, (c) white image.* 

#### 7.1.5. Super-rays for efficient Light fields processing

Participants: Matthieu Hog, Christine Guillemot.

Light field acquisition devices allow capturing scenes with unmatched post-processing possibilities. However, the huge amount of high dimensional data poses challenging problems to light field processing in interactive time. In order to enable light field processing with a tractable complexity, we have addressed, in collaboration with Neus Sabater (technicolor) the problem of light field over-segmentation [15]. We have introduced the concept of super-ray, which is a grouping of rays within and across views (see Fig.4), as a key component of a light field processing pipeline. The proposed approach is simple, fast, accurate, easily parallelisable, and does not need a dense depth estimation. We have demonstrated experimentally the efficiency of the proposed approach on real and synthetic datasets, for sparsely and densely sampled light fields. As super-rays capture a coarse scene geometry information, we have also shown how they can be used for real time light field segmentation and correcting refocusing angular aliasing.



Figure 4. Super-rays for the sparsely sampled light field in the Tsukuba dataset.

#### 7.2. Representation and compression of large volumes of visual data

Sparse representations, data dimensionality reduction, compression, scalability, perceptual coding, ratedistortion theory

#### 7.2.1. Cloud-based image and video compression

Participants: Jean Begaint, Christine Guillemot.

The emergence of cloud applications and web services has led to an increasing use of online resources for storing and exchanging images and videos. Billions of images are already stored in the cloud, and hundreds of millions are uploaded every day. Redundancy between images stored in the cloud can be leveraged to efficiently compress images by exploiting inter-images correlations. We have developed a region-based prediction scheme to exploit correlation between images in the cloud. In order to compensate the deformations between correlated images, the reference image of the cloud is first segmented into multiple regions determined from matched local features and aggregated super-pixels. We then estimate a photometric and geometric deformation model between the matched regions in the reference frame and frame to be coded. Multiple references are then generated, by applying the estimated deformation models to the reference frame, and organized in a pseudo-sequence to be differentially encoded with classic video coding tools. Experimental

results demonstrate that the proposed approach yields significant rate-distortion performance improvements compared to current coding solutions such as HEVC.

#### 7.2.2. Rate-distortion optimized tone curves for HDR video compression

Participants: David Gommelet, Christine Guillemot, Aline Roumy.

High Dynamic Range (HDR) images contain more intensity levels than traditional image formats. Instead of 8 or 10 bit integers, floating point values requiring much higher precision are used to represent the pixel data. These data thus need specific compression algorithms. The goal of the collaboration with Ericsson is to develop novel compression algorithms that allow compatibility with the existing Low Dynamic Range (LDR) broadcast architecture in terms of display, compression algorithm and datarate, while delivering full HDR data to the users equipped with HDR display. In 2016, a scalable video compression was developed offering a base layer that corresponds to the LDR data and an enhancement layer, which together with the base layer corresponds to the HDR data. In 2017 instead, we developed a backward compatible compression algorithm of HDR images, where only the LDR data are sent [14]. The novelty of the approach relies on the optimization of an invertible mapping called Tone Mapping Operator (TMO) that maps efficiently the HDR data to the LDR data. Two optimizations have been carried out in a rate-distortion sense: in the first problem, the distortion of the HDR data is minimized under the constraint of minimum LDR datarate, while in the second problem, a new constraint is added in the optimization problem to insure that LDR data are closed to some "aesthetic" a priori. Taking into account the aesthetic of the scene in video compression is indeed novel, since video compression is traditionally optimized to deliver the smallest distortion with the input data at the minimum datarate. Moreover, we provided new statistical models for estimating the distortions and the rate and showed their accuracy to the real data. Finally, a novel axis is currently carried out to efficiently exploit the temporal redundancy in HDR videos.

#### 7.2.3. Sparse image representation and deep learning for compression

Participants: Thierry Dumas, Christine Guillemot, Aline Roumy.

Deep learning is a novel research area that attempts to extract high level abstractions from data by using a graph with multiple layers. One could therefore expect that deep learning might allow efficient image compression based on these high level features. However, there are many issues that make the learning task difficult in the context of image compression. First, learning a transform is equivalent to learning an autoencoder, which is of its essence unsupervised and therefore more difficult that classical supervised learning, where deep learning has shown tremendous results. Second, the learning has to be performed under a rate-distortion criterion, and not only a distortion criterion, as is classically done in machine learning. Last but not least, deep learning, as classical machine learning, consists in two phases: (i) build a graph that can make a good representation of the data (i.e. find an architecture usually made with neural nets), and (ii) learn the parameters of this architecture from large-scale data. As a consequence, neural nets are well suited for a specific task (text or image recognition) and require one training per task. The difficulty to apply machine learning approach to image compression is that it is important to deal with a large variety of patches, and with also various compression rates. Different architectures have been proposed to design a single neural network that can work efficiently at any coding rate either by a Winner Take all approach [28] or an adaptation to the quantization noise during the training [40].

#### 7.2.4. Graph-based multi-view video representation

Participants: Christine Guillemot, Thomas Maugey, Mira Rizkallah, Xin Su.

One of the main open questions in multiview data processing is the design of representation methods for multiview data, where the challenge is to describe the scene content in a compact form that is robust to lossy data compression. Many approaches have been studied in the literature, such as the multiview and multiview plus depth formats, point clouds or mesh-based techniques. All these representations contain two types of data: i) the color or luminance information, which is classically described by 2D images; ii) the geometry information that describes the scene 3D characteristics, represented by 3D coordinates, depth maps or disparity vectors. Effective representation, coding and processing of multiview data partly rely on a proper

representation of the geometry information. The multiview plus depth (MVD) format has become very popular in recent years for 3D data representation. However, this format induces very large volumes of data, hence the need for efficient compression schemes. On the other hand, lossy compression of depth information in general leads to annoying rendering artefacts especially along the contours of objects in the scene. Instead of lossy compression of depth maps, we consider the lossless transmission of a geometry representation that captures only the information needed for the required view reconstructions. Our goal is to transmit "just enough" geometry information for accurate representation of a given set of views, and hence better control the effect of geometry lossy compression.

In 2016, we have developed a graph-based representation for complex camera configurations. In particular, a generalized Graph-Based Representation has beend eveloped which handles two views with complex translations and rotations between them. The proposed approach uses the epipolar segments to have a row-wise description of the geometry that is as simple as for rectified views. In 2017, the Graph-based Representation has been extended to build a rate-distortion optimized description of the geometry of multi-view images [22]. This work brings two major novelties. First the graph can now handle multiple views (more than 2) thanks to a recursive construction of the geometry across the views. Second, the number of edges describing the geometry information is carefully chosen with respect to a rate-distortion criterion evaluated on the reconstructed views.

An adaptation of the graph-based representations (GBR) has been proposed to describe color and geometry information of light fields (LF) in [38]. Graph connections describing scene geometry capture inter-view dependencies. They are used as the support of a weighted Graph Fourier Transform (wGFT) to encode disoccluded pixels. The quality of the LF reconstructed from the graph is enhanced by adding extra color information to the representation for a sub-set of sub-aperture images. Experiments show that the proposed scheme yields rate-distortion gains compared with HEVC based compression (directly compressing the LF as a video sequence by HEVC).

#### 7.2.5. Light fields compression using sparse reconstruction

Participants: Fatma Hawary, Christine Guillemot.

Light field data exhibits large amount of information, which poses challenging problems in terms of storage capacity, hence the need for efficient compression schemes. In collaboration with Technicolor (Dominique Thoreau and Guillaume Boisson), we have developed a scalable coding method for the light field data based on the sparsity of light fields in the angular (view) domain. A selected set of the light field sub-aperture images is encoded as a video sequence in a base layer and transmitted to the decoder. The remaining light field views are then reconstructed from the decoded subset of views, by exploiting the light field sparsity in the angular continuous Fourier domain. The reconstructed light field is enhanced using a patch-based restoration method which further exploits the light field angular redundancy.

#### 7.2.6. Light fields dimensionality reduction and compression

Participants: Elian Dib, Christine Guillemot, Xiaoran Jiang, Mikael Le Pendu.

We have investigated low rank approximation methods exploiting data geometry for dimensionality reduction of light fields. We have developed an approximation method in which homographies and the rank approximation model are jointly optimized [32]. The homographies are searched in order to align linearly correlated sub-aperture images in such a way that the batch of views can be approximated by a low rank model. The light field views are aligned using either one global homography or multiple homographies depending on how much the disparity across views varies from one depth plane to the other. The rank constraint is expressed as a product of two matrices, where one matrix contains basis vectors and where the other one contains weighting coefficients. The basis vectors and weighting coefficients can be compressed separately exploiting their respective characteristics. The optimization hence proceeds by iteratively searching for the homographies and the factored model of the input set of sub-aperture images (views), which will minimize the approximation error.

A light field compression algorithm based on a low rank approximation exploiting scene and data geometry has then be developed [18]. The best pair of key parameters (approximation rank and quantization step size), in terms of rate-distortion performance, of the algorithm are predicted based on a model learned from a set of training light fields. The model is learned as a function of several input light field features: disparity indicators defined as a function of the decay rate of the SVD values of the original and registered view matrices, as well as texture indicators defined in terms of the decay rate of SVD values computed on the central view. The parameter prediction problem is cast as a multi-output classification problem solved using a Decision Tree ensemble method, namely the Random Forest method. The approximation method is currently being extended to local super-ray based low rank models.

#### 7.3. Rendering, inpainting and super-resolution

image-based rendering, inpainting, view synthesis, super-resolution

#### 7.3.1. Transformation of the Beta distribution for color transfer

Participants: Hristina Hristova, Olivier Le Meur.

After having investigated the use of multivariate generalized Gaussian distribution in color transfer, we propose a novel transformation between two Beta distributions. The key point is that performing a Gaussianbased transformation between bounded distributions may result in out-of-range values. Furthermore, as a symmetrical distribution, the Gaussian distribution cannot model asymmetric distributions. This reveals important limitations of the Gaussian model when applied to image processing tasks and, in particular, to color transfer. To tackle these limitations of the Gaussian-based transformations, we investigate the use of bounded distributions, and more specifically, the Beta distribution. The Beta distribution is a bounded two-parameter dependent distribution, which can admit different shapes and thus, fit various data, bounded in a discrete interval. Adopting the Beta distribution to model color and light distributions of images is our key idea and motivation. The proposed transformation progressively and accurately reshapes an input Beta distribution into a target Beta distribution using four intermediate statistical transformations. Experiments have shown that the proposed method obtains more natural and less saturated results than results of recent state-of-the-art color transfer methods. Moreover, the results portray better both the target color palette and the target contrast.

#### 7.3.2. Light field inpainting and edit propagation

Participants: Oriel Frigo, Christine Guillemot, Mikael Le Pendu.

With the increasing popularity of computational photography brought by light field, simple and intuitive editing of light field images is becoming a feature of high interest for users. Light field editing can be combined with the traditional refocusing feature, allowing a user to include or remove objects from the scene, change its color, its contrast or other features.

A simple approach for editing a light field image can be obtained with an edit propagation, where first a particular subaperture view is edited (most likely the center one) and then a coherent propagation of this edit is performed through the other views. This problem is particularly challenging for the task of inpainting, as the disparity field is unknown under the occludding mask.

We have developed two methods which exploit two different light field priors, namely a low rank prior and a smoothness prior in epipolar plane images (EPI) to propagate a central view inpainting or edit to all the other views. In the first method, a set of warped versions of the inpainted central view with random homographies are vectorized and concatenated columnwise into a matrix together with the views of the light field to be inpainted. Because of the redundancy between the views, the matrix satisfies a low rank assumption enabling us to fill the region to inpaint with low rank matrix completion. To this end, a new matrix completion algorithm, better suited to the inpainting application than existing methods, has also been developed. In its simple form, our method does not require any depth prior, unlike most existing light field inpainting algorithms. The method has then been extended to better handle the case where the area to inpaint contains depth discontinuities.
In the second approach, the problem of propagating an edit from a single view to the remaining light field is solved by a structure tensor driven diffusion on the epipolar plane images [29]. Since EPIs are piecewise smooth and have no complex texture content, tensor driven diffusion is naturally suited for inpainting the EPIs as an efficient technique to obtain a coherent edit propagation. The proposed method has been shown to be useful for two applications: light field inpainting and recolorization. While the light field recolorization is obtained with a straightforward diffusion, the inpainting application is particularly challenging, as the structure tensors accounting for disparities are unknown under the occluding mask. This issue has been addressed with a disparity inpainting by means of an interpolation constrained by superpixel boundaries.



Figure 5. Overview of the proposed method. On the left, a light field with an inpainted central view and after a first edit propagation, which is performed through the center column of views (red arrows). Remaining edit propagations are performed row-by-row (blue arrows). The remaining boxes illustrate the steps of epipolar plane diffusion: Structure tensor computation where we obtain dominant structure tensors which are then spatially regularized and inpainted to estimate the structre tensors in the unknown part of the light field. Finally a tensor driven diffusion is performed on the EPIs.

#### 7.3.3. Light fields super-resolution

Participants: Christine Guillemot, Lara Younes.

Capturing high spatial resolution light fields remains technologically challenging, and the images rendered from real light fields have today a significantly lower spatial resolution compared to traditional 2D cameras. In collaboration with the University of Malta (Prof. Reuben Farrugia), we have developed an example-based super-resolution algorithm for light fields, which allows the increase of the spatial resolution of the different views in a consistent manner across all sub-aperture images of the light field [12]. To maintain consistency across all sub-aperture images of the light field, the algorithm operates on 3D stacks (called patch-volumes) of 2D-patches, extracted from the different sub-aperture images. The patches forming the 3D stack are best matches across subaperture images. A dictionary of examples is first constructed by extracting, from a training set of high- and low- resolution light fields, pairs of high- and low-resolution patch-volumes. These patchvolumes are of very high dimension. Nevertheless, they contain a lot of redundant information, hence actually lie on subspaces of lower dimension. The low- and high-resolution patch-volumes of each pair can therefore be projected on their respective low and high-resolution subspaces using e.g. Principal Component Analysis (PCA). The dictionary of pairs of projected patch-volumes (the examples) map locally the relation between the high-resolution patch volumes and their low-resolution (LR) counterparts. A linear mapping function is then learned, using Multivariate Ridge Regression (RR), between the subspaces of the low- and high- resolution patch-volumes. Each overlapping patch-volume of the low-resolution light field can then be super-resolved by a straight application of the learned mapping function (some results in Fig.6).



Figure 6. Illustration of some results on a crop of a central view with a magnification factor of 2; Top row: original (left), bicubic interpolation (right); Bottom row: super-resolved with a deep learning technique of the litterature (left), super-resolved with the proposed method (right).

This work is currently being extended on one hand by exploring how deep learning techniques can further benefit the scheme, and on the other hand by considering a hybrid system in which a 2D high resolution image, a priori not aligned with the light field views, can guide the light field super-resolution process.

#### 7.4. Distributed processing and robust communication

Information theory, stochastic modelling, robust detection, maximum likelihood estimation, generalized likelihood ratio test, error and erasure resilient coding and decoding, multiple description coding, Slepian-Wolf coding, Wyner-Ziv coding, information theory, MAC channels

#### 7.4.1. Interactive Coding for Navigation in 3D scenes (ICON 3D)

Participants: Thomas Maugey, Aline Roumy.

In order to have performing FTV systems, the data transmission has to take into account the interactivity of the user, *i.e.*, the viewpoint that is requested. In other words, a FTV system transmits to the visualisation support only what needs to be updated when a user changes its viewpoint angle (*i.e.*, the new information appearing in its vision field).

In the context of the project ICON 3D funded by the GdR-Isis, we have developed new geometry prediction algorithms for surface meshes. Given a part of a mesh, the prediction algorithm is able to estimate a neighboring mesh subset corresponding to the one newly visible after user viewpoint angle change. For each mesh of a 3D model, we have generated all the predictions possible depending on the part of the model known by the decoder. Then we have characterized the prediction error.

The question of which data representation to use for Interactive Navigation has also been studied in [20]. More precisely, the navigation domain is split in small segments, each of them coded independently. This work has developed some optimal partitioning solution for different navigation scenario.

#### 7.4.2. Correlation model selection for interactive video communication

Participants: Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy.

Interactive video communication has been recently proposed for multi-view videos. In this scheme, the server has to store the views as compactly as possible while allowing interactive navigation. Interactive navigation refers to the possibility for the user to select one view or a subset of views. To achieve this goal, the compression must be done using a model-based coding in which the correlation between the predicted view generated on the user side and the original view has to be modeled by a statistical distribution. In the context of the project Intercomm, the work published in [37] has proposed a framework for lossless fixed-length source coding to select a model among a candidate set of models that incurs the lowest extra rate cost to the system. Moreover, in cases where the depth image is available, we provide a method to estimate the correlation model.

#### 7.4.3. Optimal selection of reference sensors for spatially correlated data storage

Participants: Thomas Maugey, Aline Roumy.

Highly instrumented Smart-cities, which are now common urban policies, are facing problems of management and storage of a large volume of data coming from an increasing number of sources. In the context of the project Intercom, we have proposed a data compression method by predictive coding of spatially correlated multi-source data. In a nutshell, some sensors are selected as references. They are used to predict the other sensor values, based on a Kriging prediction. We have proposed an algorithm to optimally select both the number and the position of the reference sensors among all the ones that are stored on a server and shared with a high number of users. This work has been done in collaboration with the Inria I4S project-team, IFFSTAR and the L2S.

## 8. Bilateral Contracts and Grants with Industry

#### 8.1. Bilateral Contracts with Industry

#### 8.1.1. CIFRE contract with Envivio/ Ericsson on LDR compatible HDR video coding

Participants: Christine Guillemot, David Gommelet, Aline Roumy.

• Title : LDR-compatible coding of HDR video signals.

- Partners : Envivio.
- Funding : Cifre Envivio/Ericsson.
- Period : Oct.2014-Sept.2017.

The goal of this Cifre contract is to design solutions for LDR-compatible coding of HDR videos. This involves the study of rate-distortion optimized tone mapping operators taking into account constraints of temporal coherency to avoid the temporal flickering which results from a direct frame-by-frame application of classical tone mapping operators. The goal is also to design a coding architecture which will build upon these operators, integrating coding tools tailored to the statistics of the HDR refinement signals.

#### 8.1.2. CIFRE contract with Harmonic on image analysis for HDR video compression

Participants: Maxime Rousselot, Olivier Le Meur.

- Title : image and video analysis for HDR video compression
- Partners : Harmonic, Univ. Rennes 1
- Funding: Harmonic, ANRT
- Period: April 2016-April 2019

This project (in collaboration with Rémi Cozot, FRVSense) aims to investigate two main axes. First, we want to assess whether the representation of High Dynamic Range signal has an impact on the coding efficiency. We will focus mainly on the Hybrid Log-Gamma (HLG) and Perceptual Quantizer (PQ) OETF (Opto-Electronic Transfer Function)approaches. The former defines a nonlinear transfer function which is display-independent and able to produce high quality images without compromising the director's artistic intent. The latter approach is based on Just Noticeable Difference curve. If it turns out that this representation has an impact, the coding strategy should be adjusted with respect to the representation. In addition, specific preprocessing tools will be defined to deal with the limitations of PQ and HLG approaches.

#### 8.1.3. CIFRE contract with Technicolor on image collection analysis

Participants: Dmitry Kuzovkin, Olivier Le Meur.

- Title : Spatiotemporal retargeting and recomposition based on artistic rules
- Partners : Technicolor, Univ. Rennes 1
- Funding: Technicolor, ANRT
- Period: Nov. 2015 Nov. 2018

The goal of the project (in collaboration with Rémi Cozot, FRVSense) is to take advantage of the huge quantities of image and video data currently available - captured by both amateur and professional users - as well as the multiple copies of each scene that users often capture, to improve the aesthetic appeal of content. Additionally, given Technicolor's unique position, we propose to take advantage of insights as well as content from professional artists and colorists to learn how different content types can be enhanced.

#### 8.1.4. CIFRE contract with Technicolor on light fields editing

Participants: Christine Guillemot, Matthieu Hog.

- Title : Light fields editing
- Research axis : 7.1.5
- Partners : Technicolor, Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Oct.2015-Sept.2018.

Editing is quite common with classical imaging. Now, if we want light-fields cameras to be in the future as common as traditional cameras, this functionality should also be enabled with light-fields. The goal of the PhD is to develop methods for light-field editing, and in 2017 we have introduced the the concept of super-ray which is a grouping of rays within and across views, and developed a fast algorithm for super-ray construction(see section 7.1.5).

#### 8.1.5. CIFRE contract with Technicolor on light fields compressed representation

Participants: Christine Guillemot, Fatma Hawary.

- Title : Light fields compressed representation
- Research axis : 7.2.5
- Partners : Technicolor, Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Feb.2016-Jan.2019.

The goal of this PhD is to study reconstruction algorithms from compressed measurements based on the assumption of sparsity in the Fourier domain. The goal is to apply these algorithms to scalable compression of light fields.

#### 8.1.6. CIFRE contract with Technicolor on cloud-based image compression

Participants: Jean Begaint, Christine Guillemot.

- Title : Cloud-based image compression
- Research axis : 7.2.1
- Partners : Technicolor, Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Nov.2015-Oct.2018.

The goal of this Cifre contract is to develop a novel image compression scheme exploiting similarity between images in a cloud. The objective will therefore be to develop rate-distortion optimized affine or homographic estimation and compensation methods which will allow us to construct prediction schemes and learn adapted bases from most similar images retrieved by image descriptors. One issue to be addressed is the rate-distortion trade-off induced by the need for transmitting image descriptors.

## 9. Partnerships and Cooperations

#### 9.1. Regional Initiatives

#### 9.1.1. CominLabs/InterCom project

Participants: Aline Roumy, Thomas Maugey.

- Title : Interactive Communication (INTERCOM): Massive random access to subsets of compressed correlated data .
- Research axis : 7.4.2
- Partners : Inria-Rennes (Sirocco team and i4S team); LabSTICC, Telecom Bretagne, Signal & Communications Department; External partner: Kieffer L2S, CentraleSupelec, Univ. Paris Sud.
- Funding : Labex CominLabs.
- Period : Oct. 2016 Nov. 2019.

This project aims to develop novel compression techniques allowing massive random access to large databases. Indeed, we consider a database that is so large that, to be stored on a single server, the data have to be compressed efficiently, meaning that the redundancy/correlation between the data have to be exploited. The dataset is then stored on a server and made available to users that may want to access only a subset of the data. Such a request for a subset of the data is indeed random, since the choice of the subset is user-dependent. Finally, massive requests are made, meaning that, upon request, the server can only perform low complexity operations (such as bit extraction but no decompression/compression). Algorithms for two emerging applications of this problem will be developed: Free-viewpoint Television (FTV) and massive requests to a database collecting data from a large-scale sensor network (such as Smart Cities).

#### 9.2. European Initiatives

#### 9.2.1. FP7 & H2020 Projects

#### 9.2.1.1. ERC-CLIM

**Participants:** Pierre David, Elian Dib, Simon Evain, Christine Guillemot, Laurent Guillo, Mikael Le Pendu, Xiaoran Jiang, Jinglei Shi, Xin Su, Lara Younes.

Light fields yield a rich description of the scene ideally suited for advanced image creation capabilities from a single capture, such as simulating a capture with a different focus and a different depth of field, simulating lenses with different apertures, for creating images with different artistic intents or for producing 3D views. Light fields technology holds great promises for a number of application sectors, such as photography, augmented reality, light field microscopy, but also surveillance, to name only a few.

The goal of the ERC-CLIM project is to develop algorithms for the entire static and video light fields processing chain, going from compact sparse and low rank representations and compression to restoration, high quality rendering and editing.

#### 9.3. International Initiatives

#### 9.3.1. Inria Associate Teams Not Involved in an Inria International Labs

Title: Graph-based Omnidirectional video Processing
International Partner (Institution - Laboratory - Researcher):
Ecole Polytechnique Fédérale de Lausanne (Switzerland) - LTS4 - Pascal Frossard
Start year: 2017
See also: http://people.rennes.inria.fr/Thomas.Maugey/wp/projects/gop/

Due to new camera types, the format of the video data has become more complex than simple 2D images or videos as it was the case a few years ago. In particular, the omnidirectional cameras provide pixels on a whole sphere around a center point and enable a vision in 360°. In addition to the fact that the data size explodes with such cameras, the inherent structure of the acquired signal fundamentally differs from the 2D images, which makes the traditional video codec obsolete. In parallel of that, an important effort of research has been lead recently, especially at EPFL, to develop new processing tools for signals lying on irregular structures (graphs). It enables in particular to build efficient coding tools for new types of signals. The proposed research project will actually study how graphs can be built for defining a suitable structure on one or several 360° videos and then used for compression.

The collaboration between SIROCCO (Inria) and LTS4 (EPFL) has been very active in the recent years. However, only one-to-one collaboration was involved. When opening these new ambitious research direction, the project GOP will involve more than two or three researchers, and build a bidirectional collaboration between different people of the SIROCCO and LTS4 teams.

#### 9.3.2. Inria International Partners

#### 9.3.2.1. Informal International Partners

We have international collaborations with:

- Reuben Farrugia, Prof. at the University of Malta, with whom we continue collaborating on light field super-resolution. The collaboration strated during the sabbatical year (Sept. 2015-Aug. 2016) he spent within the team.
- Ehsan Miandji and Prof.Jonas Unger from Linkoping Univ. with whom we collaborate on compressive sampling of light fields. Ehsan Miandji has spent 1.5 month (June- July 2017) within the team.
- Chiara Galdi and Jean Luc Dugelay, prof. at Eurecom, with whom we collaborate on the application of light fields to biometry. Chiara Galdi has spent one month in the team (April 2017).
- Ole Johanssen and Prof. Bastian Goldluecke, from Univ. of Konstanz, with whom we collaborate on scene flow estimation with deep learning. Ole Johanssen has spent one month (Nov. 20- Dec. 20, 2017) in the team.
- The study on guided image inpainting is carried out in collaboration with Prof. Pascal Frossard from EPFL (Ecole Polytechique Federal de Lausanne).

## **10.** Dissemination

#### **10.1. Promoting Scientific Activities**

#### 10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Conference Program Committees

- C. Guillemot has been a member of the VISAPP 2017 international program committee.
- C. Guillemot has been a member of the international program committee of the CVPR 2017 workshop on light fields for computer vision.
- O. Le Meur has been a member of technical program committees of international conferences: QoMEX 2017, Mutual Benefits of Cognitive and Computer Vision (MBCC) 2017 workshop in conjunction with ICCV 2017, IPTA 2017, ICMR 2017.
- A. Roumy has been a member of the technical program committee of the CVPR 2017 workshop on New Trends in Image Restoration and Enhancement.
- A. Roumy has been a member of the technical program committee of the (Groupement de Recherche en Traitement du Signal et des Images) GRETSI 2017 workshop.

#### 10.1.2. Journal

#### 10.1.2.1. Member of the Editorial Boards

- C. Guillemot is senior area editor of the IEEE Trans. on Image Processing.
- C. Guillemot is associate editor of the International Journal on Mathematical Imaging and Vision.
- O. Le Meur is member of the editorial board of the IET Image Processing Journal
- A. Roumy is associate editor of the Springer Annals of Telecommunications.

#### 10.1.3. Invited Talks

- C. Guillemot gave a seminar at IST-Lisboa on Light field image processing (10th May 2017).
- C. Guillemot gave an invited talk in the context of the IEEE distinguished lecturer program, at UPM Madrid on Light field image processing (29th June 2017).
- C. Guillemot gave an invited talk in the context of the IEEE distinguished lecturer program University Autonoma de Barcelona on Light field image processing (30th June 2017).
- C. Guillemot gave an invited talk in the context of the IEEE distinguished lecturer program on "Light field image processing: analysis, representation, compression and editing", at the VISVA 2017 summer school, Singapore (8th July 2017).
- C. Guillemot gave a Keynote ("Towards dense multi-view imaging") at the Hot3D workshop jointly held with IEEE-ICME, Hong Kong, 10 July 2017.
- C. Guilemot gave a seminar at MidSweden University in light field image processing (15th Nov. 2017).
- O. Le Meur gave an invited talk at the AFIG (Association Française d'informatique Graphique). The title of the presentation was: "un nouveau regard sur l'attention visuelle".
- O. Le Meur gave an invited talk at the NEUROSTIC workshop (GDR-ISI). The title of the presentation was: "Modèles de saillance visuelle et mouvements oculaires. Un nouveau départ?".
- O. Le Meur was a captsone speaker at EuroRV3 workshop. The title of the presentation was: "The computational modelling of visual attention : saliency model vs saccadic model"
- O. Le Meur gave an invited talk at ENSTA. The title of the presentation was: "Saccadic model, a promising framework for modelling overt visual attention".
- T. Maugey gave an invited talk at EPFL (Lausanne, Switzerland) on "Massive Random Access », May 2017.
- T. Maugey gave an invited talk at the scientific day of the L2TI (University Paris XIII), December 2017.
- A. Roumy gave an introductory talk on Interactive video compression at the Comité des projets, Rennes, May 2017.
- A. Roumy gave a talk on Challenges in data compression for the IoT, at the IoTsustain Workshop, UCL, London, July 2017.

#### 10.1.4. Leadership within the Scientific Community

- C. Guillemot is member of the IEEE IVMSP technical committee
- C. Guillemot is senior member of the steering committee of IEEE Trans. on Multimedia (2016-2018).
- C. Guillemot has been appointed member of the IEEE Signal Processing Society Nominations and Appointments Committee for a two year term (2018-2019).
- A. Roumy is a member of the Executive board of the National Research group in Image and Signal Processing (GRETSI).

#### 10.1.5. Scientific Expertise

• C. Guillemot served in the HCERES panel which evaluated the CEDRIC laboratory at CNAM (Oct. 2017).

#### 10.1.6. Research Administration

- C. Guillemot is a member of Inria's evaluation committee.
- C. Guillemot is member of the "bureau du Comité des Projets".
- A. Roumy served as a member of Board of Examiners (Comité de sélection) for an Associate Professor position (Maitre de Conférences) at ENSSAT-IRISA, Lannion (61MCF1758).
- A. Roumy served as a member of Board of Examiners (Comité de sélection) for an Associate Professor position (Maitre de Conférences) in Digital Communication at CentraleSupelec, Rennes.
- A. Roumy served as a jury member for the selection of Inria CR (researcher) candidates, March and May 2017, Inria, Rennes, France
- A. Roumy is a member of the technological development and technology transfer committee at Inria Rennes
- A. Roumy was a member of Inria's PhD grant committee (CORDI-S).

### 10.2. Teaching - Supervision - Juries

#### 10.2.1. Teaching

- Master: C. Guillemot, Image and video compression, 10 hours, M2 SISEA, Univ. of Rennes 1, France.
- Undergraduate: L. Guillo, "Scientific programming ", 20 hours, University Rennes 1, "Physique, chimie et structure de la matière (PCSTM), 1st year", France.
- Master: O. Le Meur, Selective visual attention, 6 hours, M2, Univ. of Paris 8, France.
- Master: O. Le Meur, Acquisition/Image Processing/Compression, 22 hours, M2 MITIC, Univ. of Rennes 1, France.
- Engineer degree: O. Le Meur, Image Processing, video analysis and compression, 54 hours, ESIR2, Univ. of Rennes 1, France.
- Engineer degree: O. Le Meur, Visual communication, 65 hours, ESIR3, Univ. of Rennes 1, France.
- Professional training: O. Le Meur, Image Processing and OpenCV, 42 hours, Technicolor Rennes.
- Master: T. Maugey, course on 3D models in a module on "advanced video", 4 hours, M2 SISEA, Univ. of Rennes 1, France.
- Engineering degree: A. Roumy, Sparse methods in Image and Signal processing, 13 hours, INSA Rennes, 5th year, Mathematical engineering, France.
- Engineering degree: A. Roumy, Image processing, 28 hours, ECAM Rennes, 4th year, France.
- Master: A. Roumy, Foundations of Smart Sensing, 18 hours, ENSAI Bruz, Master of Science in Statistics for Smart Data, France.
- Master: A. Roumy, High dimensional statistical learning, 9 hours, University Rennes 1, Research in Computer Science (SIF) master, France.
- Master: A. Roumy, Information Theory, 15 hours, University Rennes 1, Research in Computer Science (SIF) master, France.

#### 10.2.2. Juries

- C. Guillemot has been member (rapporteur) of the PhD committee of:
  - A. Purica, Telecom ParisTech, June 2017
  - G. Nieto, Univ. Grenoble-Alpes, Oct. 2017
- C. Guillemot has been member of the PhD jury of:

- A. Aldahdooh, Univ. Nantes, Aug. 2017
- C. Peyrard, Univ. Lyon, Sept. 2017
- S. Puri, Univ. Nantes, Nov. 2017
- B. Pasdeloup, IMT-Atlantique, Dec. 2017
- O. Le Meur has been member (rapporteur) of the jury of the PhD committee of:
  - Benoit Arbelot, Univ. Grenoble Alpes, Mar. 2017
  - Celine Craye, Univ. Paris-Saclay, Apr. 2017
  - Yashas Rai, Univ. of Nantes, Aug. 2017
- Thomas Maugey has been member of the jury of the PhD committee of:
  - A. Dricot TelecomParisTech, Mar. 2017
- A. Roumy has been member of the PhD jury of:
  - V. Gouldieff, CentraleSupelec, Rennes, October 2017.

## 11. Bibliography

#### Major publications by the team in recent years

- V. CHAPPELIER, C. GUILLEMOT. Oriented wavelet transform for image compression and denoising, in "IEEE Transactions on Image Processing", 2006, vol. 15, n<sup>o</sup> 10, p. 2892-2903, http://hal.inria.fr/inria-00504227.
- [2] T. COLLEU, S. PATEUX, L. MORIN, C. LABIT. A polygon soup representation for multiview coding, in "Journal of Visual Communication and Image Representation", Feb 2010, p. 1–32.
- [3] G. COLUCCIA, A. ROUMY, E. MAGLI.Operational Rate-Distortion Performance of Single-source and Distributed Compressed Sensing, in "Communications, IEEE Transactions on", June 2014, vol. 62, n<sup>o</sup> 6, p. 2022 - 2033, IEEE Transactions on Communications [DOI: 10.1109/TCOMM.2014.2316176], https://hal.inria. fr/hal-00996698.
- [4] C. GUILLEMOT, O. LE MEUR.*Image inpainting: Overview and recent advances*, in "IEEE Signal Processing Magazine", Jan. 2014, vol. 31, n<sup>o</sup> 1, p. 127-144.
- [5] C. GUILLEMOT, A. ROUMY. Toward constructive Slepian-Wolf coding schemes, in "Distributed Source Coding: Theory, Algorithms, and Applications", P. L. DRAGOTTI, M. GASTPAR (editors), Elsevier, January 2009, https://hal.inria.fr/hal-01677745.
- [6] H. JÉGOU, C. GUILLEMOT. Robust multiplexed codes for compression of heterogeneous data, in "IEEE Transactions on Information Theory", April 2005, vol. 51, n<sup>o</sup> 4, p. 1393 - 1407, http://hal.inria.fr/inria-00604036.
- [7] O. LE MEUR, P. LE CALLET, D. BARBA, D. THOREAU.A coherent computational approach to model the bottom-up visual attention, in "IEEE Trans. On PAMI", May 2006, vol. 28, n<sup>o</sup> 5, p. 802-817.
- [8] T. MAUGEY, A. ORTEGA, P. FROSSARD.Graph-based representation for multiview image geometry, in "IEEE Transactions on Image Processing", May 2015, vol. 24, n<sup>o</sup> 5, p. 1573-1586 [DOI: 10.1109/TIP.2015.2400817], https://hal.inria.fr/hal-01116211.

[9] J. ZEPEDA, C. GUILLEMOT, E. KIJAK.Image compression using sparse representations and the iterationtuned and aligned dictionary, in "IEEE Journal on Selected Topics in Signal Processing", Sep. 2011, vol. 5, p. 1061-1073.

#### **Publications of the year**

#### **Articles in International Peer-Reviewed Journal**

- [10] M. ALAIN, C. GUILLEMOT, D. THOREAU, P. GUILLOTEL.Scalable Image Coding based on Epitomes, in "IEEE Transactions on Image Processing", May 2017, vol. 26, n<sup>o</sup> 8, p. 3624-3635 [DOI: 10.1109/TIP.2017.2702396], https://hal.inria.fr/hal-01591504.
- [11] P. BUYSSENS, O. LE MEUR, M. DAISY, D. TSCHUMPERLÉ, O. LÉZORAY. Depth-guided disocclusion inpainting of synthesized RGB-D images, in "IEEE Transactions on Image Processing", 2017, vol. 26, n<sup>o</sup> 2, p. 525-538 [DOI: 10.1109/TIP.2016.2619263], https://hal.archives-ouvertes.fr/hal-01391065.
- [12] R. A. FARRUGIA, C. GALEA, C. GUILLEMOT. Super Resolution of Light Field Images using Linear Subspace Projection of Patch-Volumes, in "IEEE Journal of Selected Topics in Signal Processing", October 2017, p. 1-14 [DOI: 10.1109/JSTSP.2017.2747127], https://hal.archives-ouvertes.fr/hal-01591488.
- [13] R. A. FARRUGIA, C. GUILLEMOT. Face Hallucination Using Linear Models of Coupled Sparse Support, in "IEEE Transactions on Image Processing", 2017, vol. 26, n<sup>o</sup> 9, p. 4562-4577 [DOI: 10.1109/TIP.2017.2717181], https://hal.archives-ouvertes.fr/hal-01591517.
- [14] D. GOMMELET, A. ROUMY, C. GUILLEMOT, M. ROPERT, J. LE TANOU. Gradient-Based Tone Mapping for Rate-Distortion Optimized Backward-Compatible High Dynamic Range Compression, in "IEEE Transactions on Image Processing", August 2017 [DOI: 10.1109/TIP.2017.2740159], https://hal.inria.fr/hal-01590258.
- [15] M. HOG, N. SABATER, C. GUILLEMOT. Super-rays for Efficient Light Field Processing, in "IEEE Journal of Selected Topics in Signal Processing", 2017 [DOI: 10.1109/JSTSP.2017.2738619], https://hal.archivesouvertes.fr/hal-01407852.
- [16] H. HRISTOVA, O. LE MEUR, R. COZOT, K. BOUATOUCH.*High-dynamic-range image recovery from flash and non-flash image pairs*, in "Visual Computer", June 2017, vol. 33, n<sup>o</sup> 6-8, p. 725 735 [DOI: 10.1007/s00371-017-1399-0], https://hal.inria.fr/hal-01650464.
- [17] H. HRISTOVA, O. LE MEUR, R. COZOT, K. BOUATOUCH. Transformation of the Multivariate Generalized Gaussian Distribution for Image Editing, in "IEEE Transactions on Visualization and Computer Graphics", November 2017, 15 [DOI: 10.1109/TVCG.2017.2769050], https://hal.inria.fr/hal-01650368.
- [18] X. JIANG, M. LE PENDU, R. A. FARRUGIA, C. GUILLEMOT.Light Field Compression with Homographybased Low Rank Approximation, in "IEEE Journal of Selected Topics in Signal Processing", 2017 [DOI: 10.1109/JSTSP.2017.2747078], https://hal.archives-ouvertes.fr/hal-01591349.
- [19] O. LE MEUR, A. COUTROT, Z. LIU, P. RÄMÄ, A. LE ROCH, A. HELO. Visual Attention Saccadic Models Learn to Emulate Gaze Patterns From Childhood to Adulthood, in "IEEE Transactions on Image Processing", October 2017, vol. 26, n<sup>o</sup> 10, p. 4777 - 4789 [DOI: 10.1109/TIP.2017.2722238], https://hal.inria.fr/hal-01650322.

- [20] R. MA, T. MAUGEY, P. FROSSARD. *Optimized Data Representation for Interactive Multiview Navigation*, in "IEEE Transactions on Multimedia", 2017, p. 1-15, https://hal.inria.fr/hal-01633660.
- [21] H. SONG, Z. LIU, H. DU, G. SUN, O. LE MEUR, T. REN. Depth-Aware Salient Object Detection and Segmentation via Multiscale Discriminative Saliency Fusion and Bootstrap Learning, in "IEEE Transactions on Image Processing", September 2017, vol. 26, n<sup>o</sup> 9, p. 4204 - 4216 [DOI: 10.1109/TIP.2017.2711277], https://hal.inria.fr/hal-01650409.
- [22] X. SU, T. MAUGEY, C. GUILLEMOT.Rate-Distortion Optimized Graph-Based Representation for Multiview Images with Complex Camera Configurations, in "IEEE Transactions on Image Processing", 2017 [DOI: 10.1109/TIP.2017.2685340], https://hal.archives-ouvertes.fr/hal-01492850.
- [23] E. VIDAL, N. STURMEL, C. GUILLEMOT, P. CORLAY, F.-X. COUDOUX.New adaptive filters as perceptual preprocessing for rate-quality performance optimization of video coding, in "Signal Processing: Image Communication", 2017, vol. 52, p. 124 - 137 [DOI: 10.1016/J.IMAGE.2016.12.003], https://hal.archivesouvertes.fr/hal-01591984.
- [24] C. VERLEYSEN, T. MAUGEY, C. DE VLEESCHOUWER, P. FROSSARD.Wide baseline image-based rendering based on shape prior regularisation, in "IEEE Transactions on Image Processing", 2017, https:// hal.inria.fr/hal-01575559.

#### **Invited Conferences**

 [25] T. MAUGEY, O. LE MEUR, Z. LIU. Saliency-based navigation in omnidirectional image, in "MMSP 2017

 IEEE International Workshop on Multimedia Signal Processing", Luton, United Kingdom, October 2017, https://hal.inria.fr/hal-01575566.

#### **International Conferences with Proceedings**

- [26] G. COLUCCIA, A. ROUMY, E. MAGLI.*Mismatched sparse denoiser requires overestimating the support length*, in "IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)", New Orleans, United States, March 2017 [DOI: 10.1109/ICASSP.2017.7953058], https://hal.inria.fr/hal-01589633.
- [27] Best Paper

P. DAVID, M. LE PENDU, C. GUILLEMOT. *White Lenslet Image Guided Demosaicing for Plenoptic Cameras*, in "MMSP 2017 - IEEE 19th International Workshop on Multimedia Signal Processing", Luton, United Kingdom, Multimedia Signal Processing (MMSP), 2017 IEEE 19th International Workshop on, October 2017, https://hal.archives-ouvertes.fr/hal-01590345.

- [28] T. DUMAS, A. ROUMY, C. GUILLEMOT.*Image Compression with Stochastic Winner-Take-All Auto-Encoder*, in "2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017)", New Orleans, United States, March 2017, https://hal.archives-ouvertes.fr/hal-01493137.
- [29] O. FRIGO, C. GUILLEMOT. Epipolar Plane Diffusion: An Efficient Approach for Light Field Editing, in "British Machine Vision Conference (BMVC)", London, France, September 2017, https://hal.archivesouvertes.fr/hal-01591525.

- [30] F. HAWARY, C. GUILLEMOT, D. THOREAU, G. BOISSON. Scalable Light Field Compression Scheme Using Sparse Reconstruction and Restoration, in "ICIP 2017 - IEEE International Conference on Image Processing", Beijing, China, September 2017, https://hal.archives-ouvertes.fr/hal-01597536.
- [31] H. HRISTOVA, O. LE MEUR, R. COZOT, K. BOUATOUCH. Perceptual metric for color transfer methods, in "IEEE International Conference on Image Processing ICIP 2017", Shanghai, China, September 2017, https://hal.inria.fr/hal-01651141.
- [32] X. JIANG, M. LE PENDU, R. A. FARRUGIA, S. S. HEMAMI, C. GUILLEMOT. Homography-Based Low Rank Approximation of Light Fields for Compression, in "IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)", New Orleans, United States, March 2017 [DOI: 10.1109/ICASSP.2017.7952369], https://hal.archives-ouvertes.fr/hal-01591315.
- [33] X. JIANG, M. LE PENDU, C. GUILLEMOT.Light field compression using depth image based view synthesis, in "IEEE International Conference on Multimedia & Expo Workshops (ICMEW)", Hong Kong, China, July 2017 [DOI: 10.1109/ICMEW.2017.8026313], https://hal.archives-ouvertes.fr/hal-01591329.
- [34] D. KUZOVKIN, T. POULI, R. COZOT, O. LE MEUR, J. KERVEC, K. BOUATOUCH. Context-aware clustering and assessment of photo collections, in "Proceedings of the symposium on Computational Aesthetics", Los Angeles, United States, July 2017 [DOI: 10.1145/3092912.3092916], https://hal.inria.fr/hal-01651171.
- [35] O. LE MEUR, A. COUTROT, A. LE ROCH, A. HELO, P. RÄMÄ, Z. LIU. Age-dependent saccadic models for predicting eye movements, in "IEEE International Conference on Image Processing - ICIP 2017", Shanghai, China, September 2017, https://hal.inria.fr/hal-01651151.
- [36] O. LE MEUR, A. COUTROT, Z. LIU, P. RÄMÄ, A. LE ROCH, A. HELO. Your gaze betrays your age, in "25th European Signal Processing Conference - EUSIPCO", Kos, Greece, August 2017, https://hal.inria.fr/ hal-01651125.
- [37] N. MAHMOUDIAN BIDGOLI, T. MAUGEY, A. ROUMY.*Correlation model selection for interactive video communication*, in "ICIP 2017 IEEE International Conference on Image Processing", Beijing, China, September 2017, https://hal.inria.fr/hal-01590627.
- [38] X. SU, M. RIZKALLAH, T. MAUGEY, C. GUILLEMOT. Graph-based light fields representation and coding using geometry information, in "IEEE International Conference on Image Processing (ICIP)", Beijing, China, IEEE, September 2017, https://hal.inria.fr/hal-01589626.

#### **National Conferences with Proceeding**

- [39] A. CRINIÈRE, A. ROUMY, T. MAUGEY, M. KIEFFER, J. DUMOULIN. Sélection optimale de capteurs de référence pour le stockage de données spatialement corrélées, in "GRETSI 2017 - XXVIème Colloque", Juan-les-Pins, France, GRETSI 2017, September 2017, https://hal.inria.fr/hal-01575610.
- [40] T. DUMAS, A. ROUMY, C. GUILLEMOT.Auto-encodeur optimisé au sens débit-distorsion : indépendant de la quantification?, in "GRETSI 2017", Juan-les-Pins, France, September 2017, https://hal.archives-ouvertes. fr/hal-01579257.

[41] T. MAUGEY, C. LE CAM, L. GUILLO. Télévision à point de vue libre et système de capture à plusieurs caméra omnidirectionnelles, in "GRETSI 2017", Juan-les-Pins, France, September 2017, https://hal.inria.fr/ hal-01575606.

#### **Other Publications**

- [42] J. BÉGAINT, D. THOREAU, P. GUILLOTEL, C. GUILLEMOT. *Region-based prediction for image compression in the cloud*, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01662639.
- [43] M. HOG, N. SABATER, C. GUILLEMOT. Dynamic Super-Rays for Efficient Light Field Video Processing, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01649342.

# **Project-Team SUMO**

# SUpervision of large MOdular and distributed systems

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: CNRS

**Université Rennes 1** 

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Proofs and Verification

## **Table of contents**

1.	Personnel	881
2.	Overall Objectives	882
	2.1.1. Necessity of quantitative models.	882
	2.1.2. Specificities of distributed systems.	882
•	2.1.3. New issues raised by large systems.	883
3.	Research Program	883
	3.1. Analysis and verification of quantitative systems	883
	3.2. Control of quantitative systems	883
	3.3. Management of large or distributed systems	884
	3.4. Data driven systems	884
4.	Application Domains	. 885
	4.1. Siliar transportation systems	005
	4.2. Collaborative workflows	005
	4.5. Conductative worknows	886
5	Highlights of the Vear	886
5.	5.1.1 New nartnershin	886
	5.1.2 Awards	887
6.	New Software and Platforms	. 887
	6.1. Active Workspaces	887
	6.2. DAXML	887
	6.3. Sigali	887
	6.4. SIMSTORS	888
	6.5. Tipex	888
7.	New Results	. 888
	7.1. Analysis and Verification of Quantitative Systems	888
	7.1.1. Diagnosability	888
	7.1.1.1. Diagnosability of repairable faults.	889
	7.1.1.2. Diagnosability degree of stochastic systems.	889
	7.1.1.3. The cost of diagnosis.	889
	7.1.2. Analysis of timed systems	889
	7.1.2.1. Determinizing timed automata.	889
	7.1.2.2. Concurrent Timed Systems.	889
	7.2. Control of Quantitative Systems	890
	7.2.1. Expressing and verifying properties of multi-agent systems	890
	7.2.1.1. Admissible strategies in controller synthesis.	890
	7.2.1.2. Strategy dependences in Strategy Logic.	890
	7.2.2. Active diagnosis	891
	7.2.2.1. Diagnosis and control of the degradation of probabilistic systems.	891
	7.2.2.2. Probabilistic Disclosure: Maximisation vs. Minimisation.	891
	7.2.3. Control and emolecement for quantitative systems	801
	7.2.3.1. Quantative determinacy and Decidability of Stochastic Games with Signals.	802
	7.2.3.2. Runtime enforcement	892
	7 2 3 4 Control of logico-numerical systems	893
	7.2.4 Smart regulation for urban trains	893
	7.3. Management of Large Distributed Systems	894
	7.3.1. Analysis and synthesis of distributed systems	894
	7.3.1.1. Control of Distributed Systems.	894

	7.3.1.2. Verification of distributed applications	894
	7.3.2. Analysis of parameterized systems	894
	7.3.2.1. Parameterized Verification of a time-synchronization protocol.	894
	7.3.2.2. Controlling population models.	894
	7.3.2.3. Handling large biological systems.	895
	7.4. Data-Driven Systems	895
	7.4.1. Incremental process discovery using Petri-net synthesis.	895
	7.4.2. An artifact model with imprecision and uncertainty	895
8.	Bilateral Contracts and Grants with Industry	<b>896</b>
	8.1.1. ADR Softwarization of Everything	896
	8.1.2. Alstom P22	896
9.	Partnerships and Cooperations	896
	9.1. National Initiatives	896
	9.1.1. ANR STOCH-MC: Model-Checking of Stochastic Systems using approximated	i algo-
	rithms (2014-2018)	896
	9.1.2. ANR HeadWork: Human-Centric Data-oriented WORKflows (2016-2020)	896
	9.1.3. IPL HAC-SPECIS: High-performance Application and Computers, Studying PErfor	mance
	and Correctness In Simulation (2016-2020)	897
	9.1.4. CNRS INS2I JCJC SensAs (2017)	897
	9.1.5. National informal collaborations	897
	9.2. International Initiatives	898
	9.2.1. Inria Associate Teams Not Involved in an Inria International Labs	898
	9.2.2. Inria International Partners	898
	9.3. International Research Visitors	898
	9.3.1. Visits of International Scientists	898
	9.3.2. Visits to International Teams	898
10.	Dissemination	<b>899</b>
	10.1. Promoting Scientific Activities	899
	10.1.1. Scientific Events Organisation	899
	10.1.1.1. General Chair, Scientific Chair	899
	10.1.1.2. Member of the Organizing Committees	899
	10.1.2. Scientific Events Selection	899
	10.1.2.1. Chair of Conference Program Committees	899
	10.1.2.2. Member of the Conference Program Committees	899
	10.1.3. Journal	899
	10.1.4. Invited Talks	900
	10.1.5. Leadership within the Scientific Community	900
	10.1.6. Research Administration	900
	10.2. Teaching - Supervision - Juries	900
	10.2.1. Teaching	900
	10.2.2. Supervision	901
	10.2.2.1. Defences	901
	10.2.2.2. PhD in progress	901
	10.2.2.3. Master2 internship supervision	901
	10.2.2.4. Other internship supervision	901
	10.2.3. Juries	902
	10.2.3.1. Juries of PhD defences:	902
	10.2.3.2. Other juries	902
14	10.3. Popularization	902
11.	вібнодгарпу	902

#### **Project-Team SUMO**

*Creation of the Team: 2013 January 01, updated into Project-Team: 2015 January 01* **Keywords:** 

#### **Computer Science and Digital Science:**

A1.2.2. - Supervision A1.3. - Distributed Systems A2.3.2. - Cyber-physical systems A2.4.2. - Model-checking A4.5. - Formal methods for security A6.4. - Automatic control A7.1. - Algorithms A7.2. - Logic in Computer Science A8.2. - Optimization A8.6. - Information theory A8.11. - Game Theory **Other Research Topics and Application Domains:** B1.1.3. - Cellular biology B1.1.9. - Bioinformatics B5.2.2. - Railway B6.2. - Network technologies B6.3.3. - Network Management

## 1. Personnel

#### **Research Scientists**

Éric Fabre [Team leader, Inria, Senior Researcher, HDR] Éric Badouel [Inria, Researcher, HDR] Nathalie Bertrand [Inria, Researcher, HDR] Blaise Genest [CNRS, Researcher, HDR] Loïc Hélouët [Inria, Researcher, HDR] Thierry Jéron [Inria, Senior Researcher, HDR] Hervé Marchand [Inria, Researcher, HDR] Nicolas Markey [CNRS, Senior Researcher, HDR] Ocan Sankur [CNRS, Researcher]

#### PhD Students

Hugo Bazille [Inria] Sihem Cherrared [Orange Labs] Arij Elmajed [Nokia, from Feb 2017] Bruno Karelovic [Univ. Paris VII] Abd El Karim Kecir [ALSTOM Transport] Engel Lefaucheux [Univ de Rennes I] The Anh Pham [Univ de Rennes I, from Sep 2017] Matthieu Pichené [Inria] Victor Roussanaly [Univ de Rennes I, from Sep 2017]

#### Visiting Scientists

Robert Nsaibirni [University of Yaounde, until Jun 2017] Shauna Laurene Ricker [Mount Allison Univ., Canada, from May 2017 until Jul 2017]

Administrative Assistant Laurence Dinh [Inria]

## 2. Overall Objectives

#### 2.1. Overall objectives

Most software-driven systems we commonly use in our daily life are huge hierarchical assemblings of components. This observation runs from the micro-scale (multi-core chips) to the macro-scale (data centers), and from hardware systems (telecommunication networks) to software systems (choreographies of web services). The main characteristics of these pervasive applications are size, complexity, heterogeneity, and modularity (or concurrency). Besides, several such systems are actively used before they are fully mastered, or they have grown so much that they now raise new problems that are hardly manageable by human operators. While these systems and applications are becoming more essential, or even critical, the need for their *reliability, efficiency* and *manageability* becomes a central concern in computer science. The main objective of SUMO is to develop theoretical tools to address such challenges, according to the following axes.

#### 2.1.1. Necessity of quantitative models.

Several disciplines in computer science have of course addressed some of the issues raised by large systems. For example, formal methods (essentially for verification purposes), discrete-event systems (diagnosis, control, planning, and their distributed versions), but also concurrency theory (modelling and analysis of large concurrent systems). Practical needs have oriented these methods towards the introduction of quantitative aspects, such as time, probabilities, costs, and their combinations. This approach drastically changes the nature of questions that are raised. For example, verification questions become the reachability of a state in a limited time, the average sojourn duration in a state, the probability that a run of the system satisfies some property, the existence of control strategies with a given winning probability, etc. In this setting, exact computations are not always appropriate as they may end up with unaffordable complexities, or even with undecidability. Approximation strategies then offer a promising way around, and are certainly also a key to handling large systems. Approaches based on discrete-event systems follow the same trend towards quantitative models. For diagnosis aspects, one is interested in the most likely explanations to observed malfunctions, in the identification of the most informative tests to perform, or in the optimal placement of sensors. For control problems, one is of course interested in optimal control, in minimizing communications, in the robustness of the proposed controllers, in the online optimization of QoS (Quality of Service) indicators, etc.

#### 2.1.2. Specificities of distributed systems.

While the above questions have already received partial answers, they remain largely unexplored in a distributed setting. We focus on structured systems, typically a network of dynamic systems with known interaction topology, the latter being either static or dynamic. Interactions can be synchronous or asynchronous. The state-space explosion raised by such systems has been addressed through two techniques. The first one consists in adopting true-concurrency models, which take advantage of the parallelism to reduce the size of the trajectory sets. The second one looks for modular or distributed "supervision" methods, taking the shape of a network of local supervisors, one per component. While these approaches are relatively well understood, their mixing with quantitative models remains a challenge (as an example, there exists no proper setting assembling concurrency theory with stochastic systems). This field is largely open both for modeling, analysis and verification purposes, and for distributed supervision techniques. The difficulties combine with the emergence of data-driven distributed systems (as web services or data centric systems), where the data exchanged by the various components influence both the behaviors of these components and the quantitative aspects of their reactions (e.g. QoS). Such systems call for symbolic or parametric approaches for which a theory is still missing.

#### 2.1.3. New issues raised by large systems.

Some existing distributed systems like telecommunication networks, data centers, or large-scale web applications have reached sizes and complexities that reveal new management problems. One can no longer assume that the model of the managed systems is static and fully known at any time and any scale. To scale up the management methods to such applications, one needs to be able to design reliable abstractions of parts of the systems, or to dynamically build a part of their model, following the needs of the management functions to realize. Besides, one does not wish to define management objectives at the scale of each single component, but rather to pilot these systems through high-level policies (maximizing throughput, minimizing energy consumption, etc.) These distributed systems and management problems have connections with other approaches for the management of large structured stochastic systems, such as Bayesian networks (BN) and their variants. The similarity can actually be made more formal: inference techniques for BN rely on the concept of conditional independence, which has a counterpart for networks of *dynamic* systems and is at the core of techniques like distributed diagnosis, distributed optimal planning, or the synthesis of distributed controllers. The potential of this connection is largely unexplored, but it suggests that one could derive from it good approximate management methods for large distributed dynamic systems.

## 3. Research Program

#### 3.1. Analysis and verification of quantitative systems

The overall objective of this axis is to develop the quantitative aspects of formal methods while maintaining the tractability of verification objectives and progressing toward the management of large systems. This covers the development of relevant modeling formalims, to nicely weave time, costs and probabilities with existing models for concurrency. We plan to further study time(d) Petri nets, networks of timed automata (with synchronous or asynchronous communications), stochastic automata, partially-observed Markov decision processes, etc. A second objective is to develop verification methods for such quantitative systems. This covers several aspects: quantitative verification questions (e.g. computing an optimal scheduling policy), Boolean questions on quantitative features (deciding whether some probability is greater than a threshold), robustness issues (will a system have the same behaviors if some parameter is slightly altered?), etc. Our goal is to explore the frontier between decidable and undecidable problems, or more pragmatically tractable and untractable problems. Of course, there is a tradeoff between the expressivity and the tractability of a model. Models that incorporate distributed aspects, probabilities, time, etc., are typically untractable. In such a case, abstraction or approximation techniques are a workaround that we will explore.

Here are some precise topics that we place in our agenda:

- analysis of diagnosability and opacity properties for stochastic systems;
- verification of time(d) Petri nets;
- robustness analysis for timed and/or stochastic systems;
- abstraction techniques for quantitative systems.

#### **3.2.** Control of quantitative systems

The main objective of this research axis is to explore the quantitative and/or distributed extensions of classical control problems. We envision control in its widest meaning of driving a system in order to guarantee or enforce some extra property (i.e. not guaranteed by the system alone), in a partially- or totally-observed setting. This property can either be logical (e.g. reachability or safety) or quantitative (e.g. reach some performance level). These problems have of course an offline facet (e.g. controller design, existence of a policy/strategy) and an online facet (e.g. algorithm to select some optimal action at runtime).

Our objectives comprise classical controler synthesis for discrete-event systems, with extensions to temporal/stochastic/reward settings. They also cover maintaining or maximizing extra properties such as diagnosability or opacity, for example in stochastic systems. We also target further analysis of POMDPs (partiallyobserved Markov decision processes), and multi-agent versions of policy synthesis relying on tools from game theory. We aim at adressing some control problems motivated by industrial applications, that raise issues like the optimal control of timed and stochastic discrete-event systems, with concerns like robustness to perturbations and multicriteria optimization. Finally, we also plan to work on modular testing, and on runtime enforcement techniques, in order to garantee extra logical and temporal properties to event flows.

#### 3.3. Management of large or distributed systems

The generic terms of "supervision" or "management" of distributed systems cover problems like control, diagnosis, sensor placement, planning, optimization, (state) estimation, parameter identification, testing, etc. This research axis examines how classical settings for such problems can scale up to large or distributed systems. Our work will be driven by considerations like: how to take advantage of modularity, how to design approximate management algorithms, how to design relevant abstractions to make large systems more tractable, how to deal with models of unknown size, how to design mechanisms to obtain relevant models, etc.

As more specific objectives, let us mention:

- Parametric-size systems: how to verify properties of distributed systems with an unknown number of components;
- Approximate management methods: we will explore the extension of ideas developed for Bayesian inference in large-scale stochastic systems (such as turbo-algorithms) to the field of modular dynamic systems. When component interactions are sparse, even if exact management methods are unaccessible (for diagnosis, planning, control, etc.), good approximations based on local computations may be accessible;
- Model abstraction: we will explore techniques to design more tractable abstractions of stochastic dynamic systems defined on large sets of variables;
- Self-modelling, which consists in managing large-scale systems that are known by their building rules, but where the specific instance is only discovered on-the-fly at runtime. The model of the managed system is built on-line, following the needs of the management algorithms;
- Distributed control: we will tackle issues related to asynchronous communications between local controllers, and to abstraction techniques allowing to address large systems;
- Test and enforcement: we will tackle coverage issues for the test of large systems, and the test and enforcement of properties for timed models, or for systems handling data.

#### **3.4. Data driven systems**

Data-driven systems are systems whose behaviour depends both on explicit workflows (scheduling and durations of tasks, calls to possibly distant services, ...) and on the data processed by the system (stored data, parameters of a request, results of a request, ...). This family of systems covers workflows that convey data (business processes or information systems), transactional systems (web stores), large databases managed with rules (banking systems), collaborative environments (crowds, health systems), etc. These systems are distributed, modular, and open: they integrate components and sub-services distributed over the web, and accept requests from clients. Our objective is to provide validation and supervision tools for such systems. To achieve this goal, we have to solve several challenging tasks:

• provide realistic models, and sound automated abstraction techniques, to reason on models that are reasonable abstractions of real systems. These models should be able to encompass modularity, distribution, in a context where workflows and data aspects are tightly connected;

- address design of data driven systems in a declarative way: declarative models are another way to
  handle data-driven systems. Rather than defining the explicit workflows and their effects on data,
  rule-based models state how actions are enacted in terms of the shape (pattern matching) or value
  of the current data. We think that distributed rewriting rules or attributed grammars can provide
  a practical yet formal framework for maintenance, by providing a solution to update mandatory
  documentation during the lifetime of an artifact.
- provide tractable solutions for validation of models: frequent issues are safety questions (can a system reach some bad configuration?), but also liveness (workflows progess), ... These questions should not only remain decidable on our models, but also with efficient computational methods.
- address QoS management in large reconfigurable systems: data-driven distributed systems often have constraints in terms of QoS. This QoS questions adress performance issues, but also data quality. This calls for an analysis of quantitative features and for reconfiguration techniques to meet desired QoS.

## 4. Application Domains

#### **4.1. Smart transportation systems**

The smart-city trend aims at optimizing all functions of future cities with the help of digital technologies. We focus on the segment of urban trains, which will evolve from static and scheduled offers to reactive and eventually on-demand transportation offers. We address two challenges in this field. The first one concerns the optimal design of robust subway lines. The idea is to be able to evaluate, at design time, the performance of time tables and of different regulations policies. In particular, we focus on robustness issues: how can small perturbations and incidents be accommodated by the system, how fast will return to normality occur, when does the system become unstable? The second challenge concerns the design of new robust regulation strategies to optimize delays, recovery times, and energy consumption at the scale of a full subway line. These problems involve large-scale discrete-event systems, with temporal and stochastic features, and translate into robustness assessment, stability analysis and joint numerical/combinatorial optimization problems on the trajectories of these systems.

#### 4.2. Management of telecommunication networks and of data centers

Telecommunication-network management is a rich provider of research topics for the team, and some members of SUMO have a long background of contacts and transfer with industry in this domain. Networks are typical examples of large distributed dynamic systems, and their management raises numerous problems ranging from diagnosis (or root-cause analysis), to optimization, reconfiguration, provisioning, planning, verification, etc. They also bring new challenges to the community, for example on the modeling side: building or learning a network model is a complex task, specifically because these models should reflect features like the layering, the multi-resolution view of components, the description of both functions, protocols and configuration, and they should also reflect dynamically-changing architectures. Besides modeling, management algorithms are also challenged by features like the size of systems, the need to work on abstractions, on partial models, on open systems, etc. The networking technology is now evolving toward software-defined networks, virtualized-network functions, multi-tenant systems, etc., which reinforces the need for more automation in the management of such systems.

Data centers are another example of large-scale modular dynamic and reconfigurable systems: they are composed of thousands of servers, on which virtual machines are activated, migrated, resized, etc. Their management covers issues like troubleshooting, reconfiguration, optimal control, in a setting where failures are frequent and mitigated by the performance of the management plane. We have a solid background in the coordination of the various autonomic managers that supervise the different functions/layers of such systems (hardware, middleware, web services, ...) Virtualization technologies now reach the domain of networking, and telecommunication operators/vendors evolve towards providers of distributed open clouds. This convergence of IT and networking strongly calls for new management paradigms, which is an opportunity for the team.

This application domain will be revived in the team by a collaboration with Orange Labs (1 Cifre PhD in the common lab Orange/Inria) and a collaboration with Nokia Bell Labs (1 Cifre PhD, and participation to the joint research team "Softwarization of Everything" of the common lab Nokia Bell Labs/Inria).

#### 4.3. Collaborative workflows

A current trend is to involve end-users in collection and analysis of data. Exemples of this trend are contributive science, crisis-management systems, and crowds. All these applications are data-centric and user-driven. They often are distributed and involve complex, and sometimes dynamic workflows. In many cases, there are strong interactions between data and control flows: indeed, decisons taken regarding the next tasks to be launched highly depend on collected data. For instance, in an epidemic-surveillance system, the aggregation of various reported disease cases may trigger alerts. Another example is crowds where user skills are used to complete tasks that are better performed by humans than computers. In return, this requires addressing imprecise and sometimes unreliable answers. We address several issues related to complex workflows and data. We study declarative and dynamic models that can handle workflows, data, uncertainty, and competence management.

Once these models are mature enough, we plan to experiment them on real use cases from contributive science, health-management systems, and crowd platforms using prototypes. We also plan to define abstaction schemes allowing formal reasonning on these systems.

#### 4.4. Systems Biology

Systems Biology is a recent topic in SUMO. In systems biology, many continuous variables interact together. Biological systems are thus good representatives for large complex quantitative systems, for which we are developing analysis and management methods. For instance, the biological pathway of apoptosis explains how many molecules interact inside a cell, triggered by some outside signal (drug, etc.), eventually leading to the death of the cell by apoptosis. While intrinsically quantitative in nature and in problems, data are usually noisy and problems need not be answered with ultimate precision. It thus seems reasonable to resort to approximations in order to handle the state-space explosion resulting from the high dimensionality of biological systems.

We are developing models and abstraction tools for systems biology. Studying these models suggests new reduction methods, such as considering populations instead of explicitly representing every single element into play (be it cells, molecules, etc): we thus develop algorithm handling population symbolically, either in a continuous (distributions) or a discrete (parametric) way. An intermediate goal is to speed-up analysis of such systems using abstractions, and a long term goal is to develop top-down model-checking methods that can be run on these abstractions.

## 5. Highlights of the Year

#### **5.1. Highlights of the Year**

#### 5.1.1. New partnership

Several members of the team are involved in the joint research team "Softwarization of Everything", part of the joint research lab of Nokia Bell Labs France and Inria. This activity will finance two PhDs in the team, related to the management and control of software-defined networks.

#### 5.1.2. Awards

- Engel Lefaucheux received the best young-researcher-paper award ("Prix Jeune Chercheur") at MSR 2017 for his paper titled *Diagnostic et contrôle de la dégradation des systèmes probabilistes*.
- Nicolas Markey was awarded an *Allocation d'Installation Scientifique* (at senior-researcher level) from Rennes Métropole.

BEST PAPERS AWARDS :

[42] MSR 2017 - Modélisation des Systèmes Réactifs. N. BERTRAND, S. HADDAD, E. LEFAUCHEUX.

## 6. New Software and Platforms

#### **6.1. Active Workspaces**

KEYWORDS: Active workspace - Collaborative systems - Artifact centric workflow system SCIENTIFIC DESCRIPTION: Tool for computer supported cooperative work where a user's workspace is given by an active structured repository containing the pending tasks together with information needed to perform the tasks. Communication between active workspaces is asynchronous using message passing. The tool is based on the model of guarded attribute grammars.

- Authors: Éric Badouel and Robert Nsaibirni
- Contact: Éric Badouel
- URL: http://people.rennes.inria.fr/Eric.Badouel/Research/ActiveWorkspaces.html

#### 6.2. DAXML

KEYWORDS: XML - Web Services - Distributed Software - Active documents

SCIENTIFIC DESCRIPTION: DAXML is an interpreter and implementation of Distributed Active Documents, a formalism for data centric design of Web Services. This implementation is based on a REST framework, and can run on a network of machines connected to internet and equiped with JAVA.

FUNCTIONAL DESCRIPTION: This prototype interprets distributed Active XML documents. It can be used to deploy services defined as active documents over the web.

- Participants: Benoît Masson and Loïc Hélouët
- Contact: Loïc Hélouët
- URL: http://www.irisa.fr/sumo/Software/DAXML/

#### 6.3. Sigali

FUNCTIONAL DESCRIPTION: Sigali is a model-checking tool that operates on ILTS (Implicit Labeled Transition Systems, an equational representation of an automaton), an intermediate model for discrete event systems. It offers functionalities for verification of reactive systems and discrete controller synthesis. The techniques used consist in manipulating the system of equations instead of the set of solutions, which avoids the enumeration of the state space. Each set of states is uniquely characterized by a predicate and the operations on sets can be equivalently performed on the associated predicates. Therefore, a wide spectrum of properties, such as liveness, invariance, reachability and attractivity, can be checked. Algorithms for the computation of predicates on states are also available. Sigali is connected with the Polychrony environment (Tea project-team) as well as the Matou environment (VERIMAG), thus allowing the modeling of reactive systems by means of Signal Specification or Mode Automata and the visualization of the synthesized controller by an interactive simulation of the controlled system.

• Contact: Hervé Marchand

#### 6.4. SIMSTORS

Simulator for stochastic regulated systems

KEYWORDS: Simulation - Public transport - Stochastic models - Distributed systems

FUNCTIONAL DESCRIPTION: SIMSTORS is a software for the simulation of stochastic concurrent timed systems. The heart of the software is a variant of stochastic and timed Petri nets, whose execution is controlled by a regulation policy (a controller), or a predetermined theoretical schedule. The role of the regulation policy is to control the system to realize obbjectives or a schedule when it exists with the best possible precision. SIMSTORS is well adapted to represent systems with randomness, parallelism, tasks scheduling, and resources. It is currently in use within collaboration P22 with Asltom Transport, where it is used to model metro traffic and evaluate performance of regulation solutions. This software allows for step by step simulation, but also for efficient performance analysis of systems such as production cells or train systems. The initial implementation was released in 2015, and the software is protected by the APP.

In 2017, SIMSTORS has been extended along two main axes: on one hand, SIMSTORS models were extended to handle situations where shared resources can be occupied by more than one object ( this is of paramount importance to represent conveyors, roads occupied by cars, or train tracks with smoothed scheduling allowing shared sections among trains) with priorities, constraint on their ordering and individual characteristics. This allows for instance to model vehicles with different speeds on a road, while handling safety distance constraints. On the other hand, SIMSTORS models were extended to allow control of stochastic nets based on decision rules that follow optimization schemes.

- Participants: Abd El Karim Kecir and Loïc Hélouët
- Contact: Loïc Hélouët
- URL: http://www.irisa.fr/sumo/Software/SIMSTORS/

#### 6.5. Tipex

#### TImed Properties Enforcement during eXecution

KEYWORDS: Monitoring - Controller synthesis - Formal methods

FUNCTIONAL DESCRIPTION: We are implementing a prototype tool named Tipex (TImed Properties Enforcement during eXecution) for the enforcement of timed properties. Tipex is based on the theory and algorithms that we develop for the synthesis of enforcement monitors for properties specified by timed automata (TA). The prototype is developped in python, and uses the PyUPPAAL and DBMpyuppaal libraries of the UPPAAL tool . It is currently restricted to safety and co-safety timed property. The property provided as input to the tool is a TA that can be specified using the UPPAAL tool, and is stored in XML format. The tool synthesizes an enforcement monitor from this TA, which can then be used to enforce a sequence of timed events to satisfy the property. Experiments have been conducted on a set of case studies. This allowed to validate the architecture and feasibility of enforcement monitoring in a timed setting and to have a first assessment of performance (and to what extent the overhead induced by monitoring is negligible).

- Participants: Thierry Jéron, Srinivas Pinisetty and Hervé Marchand
- Contact: Thierry Jéron

## 7. New Results

#### 7.1. Analysis and Verification of Quantitative Systems

#### 7.1.1. Diagnosability

Participants : Hugo Bazille, Éric Fabre, Blaise Genest, Loïc Hélouët, Hervé Marchand, Engel Lefaucheux

#### 7.1.1.1. Diagnosability of repairable faults.

Diagnosability (i.e., the existence of a diagnoser detecting faults in partially-observable systems) can be decided in polynomial time, relying on the so-called twin-machine construction. We have examined the case of repairable faults, and a notion of diagnosability that requires the detection of the fault before it is repaired. We have extended a contribution of 2016 to show that diagnosability of faults and of their repair could help counting the number of occurred faults. It was proved [51] that diagnosability with repair is a *PSPACE*-complete problem. We have completed this result, showing that the close notion of P-diagnosability (diagnosability of a fault even after it is repaired) is also *PSPACE*-complete [20].

#### 7.1.1.2. Diagnosability degree of stochastic systems.

For stochastic systems, several diagnosability properties have been defined. The simplest one, also called A-diagnosability, characterizes the fact that after each fault, detection will almost surely occur. We have considered quantitative versions of the problem, to determine how much a system is diagnosable (when it is not diagnosable for sure). This amounts to characterizing the probability that a faulty run will lead to detection. We have proposed several notions of dignosability degree. Their derivation is generally *NP*-hard, but we have identified situations where complexity becomes polynomial. Besides, we have developed techniques to compute the different moments of the detection delay (mean, variance and upper moments). This allows one to compare systems with similar detection degrees, but that can react faster to faults. In some cases, one may be able to tune a system and trade diagnosability degree againts a faster detection. This approach also yields the distribution of fault location (in time) once detection takes place. Given the first moments of the detection delay, one is also able to compute (sometimes tight) bounds on the response time, for example to lower bound the probability that detection takes place at most *T* seconds/events after the fault [31].

#### 7.1.1.3. The cost of diagnosis.

We addressed diagnosability and its cost for safe Petri nets. In [37] we have defined an energy-like cost model for Petri nets: transitions can consume or restore energy of the system. We then have defined a partial-order representation for state estimation, and extend the cost model and the capacities of diagnosers. Diagnosers are allowed to use additional energy to refine their estimations. Diagnosability is then seen as an energy game: checking whether disambiguation mechanisms are sufficient to allow diagnosability is in *2EXPTIME*, and one can also decide in *2EXPTIME* whether diagnosability under budget constraint holds.

#### 7.1.2. Analysis of timed systems

Participants : Nicolas Markey, Loïc Hélouët

#### 7.1.2.1. Determinizing timed automata.

In [35], we introduce a new formalism called *automata over a timed domain*, which generalizes timed automata; this formalism provides an adequate framework for determinization. In our formalism, determinization w.r.t. timed language is always possible at the cost of changing the timed domain. We give a condition for determinizability of automata over a timed domain *without changing the timed domain*, which allows us to recover several known determinizable classes of timed systems, such as strongly-non-zeno timed automata, integer-reset timed automata, perturbed timed automata, etc. Moreover, in the case of timed automata, this condition encompasses most determinizability conditions from the literature. Our aim now is to extend this work towards more efficient algorithms for monitoring timed systems.

#### 7.1.2.2. Concurrent Timed Systems.

Time Petri nets (TPNs) are a classical extension of Petri nets with timing constraints attached to transitions, for which most verification problems are undecidable. We consider TPNs under a strong semantics with multiple enablings of transitions. This year, we have extend a work started in 2016, focusing on a structural subclass of unbounded TPNs, where the underlying untimed net is free choice, and showed that it enjoys nice properties in the timed setting under a multi-server semantics [46], [25]. In particular, we have showed that the questions of firability (whether a chosen transition can fire), and termination (whether the net has a non-terminating run) are decidable for this class. Next, we have considered the problem of robustness under guard enlargement and guard shrinking, i.e., whether a given property is preserved even if the system is

implemented on an architecture with imprecise time measurement. For unbounded free choice TPNs with a multi-server semantics, we have show decidability of robustness of firability and of termination under both guard enlargement and shrinking.

#### 7.2. Control of Quantitative Systems

#### 7.2.1. Expressing and verifying properties of multi-agent systems

Participants : Ocan Sankur, Nicolas Markey

#### 7.2.1.1. Admissible strategies in controller synthesis.

In game theory, a strategy is dominated by another one if the latter systematically yields a payoff as good as the former, while also yielding a better payoff in some cases. A strategy is admissible if it is not dominated. This notion is well-studied in game theory and is useful to describe the set of strategies that are "reasonable" (i.e., whose choice can be justified; here, no players would play a dominated strategy, since better strategies exist). Recent works studied this notion in graph games with omega-regular objectives and investigated its applications in controller synthesis. For multi-agent controller synthesis, admissibility can be used as a hypothesis on the behaviors of each agent, thus enabling a compositional reasoning framework for controller synthesis.

We continue the study of admissibility in controller synthesis with three developments detailed as follows:

- In [29], we study the characterization and computation of admissible strategies in multiplayer concurrent games. We study both deterministic strategies and randomized ones with almost-sure winning criteria. We prove that admissible strategies always exist in concurrent games, and we characterise them precisely. Then, when the objectives of the players are omega-regular, we show how to perform assume-admissible synthesis, i.e., how to compute admissible strategies that win (almost surely) under the hypothesis that the other players play admissible strategies only.
- In [30], we study timed games, which are multiplayer games played on arena defined by timed automata, which are a particular case of concurrent games. First, we show that admissible strategies may not exist in timed games with a continuous semantics of time, even for safety objectives. Second, we show that the discrete time semantics of timed games is better behaved w.r.t. admissibility: the existence of admissible strategies is guaranteed in that semantics. Third, we provide symbolic algorithms to solve the model-checking problem under admissibility and the assume-admissible synthesis problem for real-time non-zero sum n-player games for safety objectives.
- In [26], we study admissible strategies in games with imperfect information. We show that in stark contrast with the perfect information variant, admissible strategies are only guaranteed to exist when players have objectives that are closed sets. As a consequence, we also study decision problems related to the existence of admissible strategies for regular games as well as finite duration games.

#### 7.2.1.2. Strategy dependences in Strategy Logic.

Strategy Logic (SL) is a very expressive logic for specifying and verifying properties of multi-agent systems: in SL, one can quantify over strategies, assign them to agents, and express properties of the resulting plays (using linear-time temporal logic). This defines a very expressive framework, encompassing e.g. (pure) Nash equilibria, or admissibility. Such a powerful framework has two drawbacks: first, SL model checking has non-elementary complexity; second, the exact semantics of SL is rather intricate, and may not correspond to what is expected.

In [49], we focus on *strategy dependences* in SL, by tracking how existentially-quantified strategies in a formula may (or may not) depend on other strategies selected in the formula. We study different kinds of dependences, refining a previous approach [52], and prove that they give rise to different satisfaction relations. In the setting where strategies may only depend on what they have observed, we identify a large fragment of SL for which we prove model checking can be performed in *2EXPTIME*.

#### 7.2.2. Active diagnosis

Participants : Nathalie Bertrand, Blaise Genest, Engel Lefaucheux

#### 7.2.2.1. Diagnosis and control of the degradation of probabilistic systems.

Active diagnosis is performed by a controler so that a system becomes diagnosable. In order to avoid the controler to degrade the functioning of the system too much, one often provides it with an additional objective specifying the desired quality of service.

In the context of probabilistic systems, a possible specification consists in requiring a positive probability of infinite correct runs, referred to as the safe active diagnosis. In [42], we introduced two alternative specifications. First  $(\gamma, v)$ -correction of a system associates with an execution a correction value which depends on a discount factor  $\gamma$ , and the controler must ensure an expected correction value greater than a threshold v. Second,  $\alpha$ -persistence requires that asymptotically, at each time unit, a proportion at least  $\alpha$  of runs that were correct so far remain correct.

Our contributions are twofold. On the one hand, from a semantical viewpoint, we make explicit the equivalences and (non-)implications between the various notions, for finite-state systems as well as infinite-state ones. On the other hand, algorithmically, we establish the decidability frontier of the corresponding decision problems, and for decidable problems characterize their precise complexity, together with algorithms to design controlers.

#### 7.2.2.2. Probabilistic Disclosure: Maximisation vs. Minimisation.

We consider opacity questions where an observation function provides to an external attacker a view of the states along executions and secret executions are those visiting some secret state from a fixed subset. Disclosure occurs when the observer can deduce from a finite observation that the execution is secret. In a probabilistic and non deterministic setting, where an internal agent can choose between actions, there are two points of view, depending on the status of this agent: the successive choices can either help the attacker trying to disclose the secret, if the system has been corrupted, or they can prevent disclosure as much as possible if these choices are part of the system design. In the former situation, corresponding to a worst case, the disclosure value is the supremum over the strategies of the probability to disclose the secret (maximisation), whereas in the latter case, the disclosure is the infimum (minimisation). We address quantitative problems (relation between the optimal value and a threshold) and qualitative ones (when the threshold is zero or one) related to both forms of disclosure for a fixed or finite horizon. For all problems, we characterise their decidability status and their complexity. Surprisingly, while in maximisation problems optimal strategies may be chosen among deterministic ones, it is not the case for minimisation problems, but more minimisation problems than maximisation ones are decidable. These results appeared in [36].

#### 7.2.3. Control and enforcement for quantitative systems

Participants : Nathalie Bertrand, Blaise Genest, Thierry Jéron, Hervé Marchand, Nicolas Markey

#### 7.2.3.1. Qualitative determinacy and Decidability of Stochastic Games with Signals.

In [17], we consider two-person zero-sum stochastic games with signals, a standard model of stochastic games with imperfect information. The only source of information for the players consists of the signals they receive; they cannot directly observe the state of the game, nor the actions played by their opponent, nor their own actions.

We are interested in the existence of almost-surely winning or positively winning strategies, under reachability, safety, Büchi, or co-Büchi winning objectives, and the computation of these strategies when the game has finitely many states and actions. We prove two qualitative determinacy results. First, in a reachability game, either player 1 can achieve almost surely the reachability objective, or player 2 can achieve surely the dual safety objective, or both players have positively winning strategies. Second, in a Büchi game, if player 1 cannot achieve almost surely the Büchi objective, then player 2 can ensure positively the dual co-Büchi objective. We prove that players only need strategies with finite memory. The number of memory states needed to win with finite-memory strategies ranges from one (corresponding to memoryless strategies) to doubly

exponential, with matching upper and lower bounds. Together with the qualitative determinacy results, we also provide fix-point algorithms for deciding which player has an almost-surely winning or a positively winning strategy and for computing an associated finite-memory strategy. Complexity ranges from *EXPTIME* to *2EXPTIME*, with matching lower bounds. Our fix-point algorithms also enjoy a better complexity in the cases where one of the players is better informed than their opponent.

Our results hold even when players do not necessarily observe their own actions. The adequate class of strategies, in this case, is mixed or general strategies (they are equivalent). Behavioral strategies are too restrictive to guarantee determinacy: it may happen that one of the players has a winning general strategy but none of them has a winning behavioral strategy. On the other hand, if a player can observe their actions, then general, mixed, and behavioral strategies are equivalent. Finite-memory strategies are sufficient for determinacy to hold, provided that randomized memory updates are allowed.

#### 7.2.3.2. Average-energy games.

In [34], we consider average-energy games, where the goal is to minimize the long-run average of the accumulated weight (seen as an *energy level*) in a two-player game on a finite-state weighted automaton. Decidability of average-energy games with a lower-bound constraint on the energy level (but no upper bound) is an open problem; in particular, there is no known upper bound on the memory that is required for winning strategies.

By reducing average-energy games with lower-bounded energy to infinite-state mean-payoff games and analyzing the frequency of low-energy configurations, we show an almost tight doubly-exponential upper bound on the necessary memory, and that the winner of average-energy games with lower-bounded energy can be determined in doubly-exponential time. We also prove *EXPSPACE*-hardness of this problem.

Finally, we consider multi-dimensional extensions of all types of average-energy games: without bounds, with only a lower bound, and with both a lower and an upper bound on the energy. We show that the fully-bounded version is the only case to remain decidable in multiple dimensions.

#### 7.2.3.3. Runtime enforcement.

The journal paper [23] details our work about predictive runtime enforcement, done in collaboration with University Aalto (Finland) and Inria CORSE/LIG Grenoble.

Runtime enforcement (RE) is a technique to ensure that the (untrustworthy) output of a black-box system satisfies some desired properties. In RE, the output of the running system, modeled as a sequence of events, is fed into an enforcer. The enforcer ensures that the sequence complies with a certain property, by delaying or modifying events if necessary. This paper deals with predictive runtime enforcement, where the system is not entirely black-box, but we know something about its behavior. This a priori knowledge about the system allows to output some events immediately, instead of delaying them until more events are observed, or even blocking them permanently. This in turn results in better enforcement policies. We also show that if we have no knowledge about the system, then the proposed enforcement mechanism reduces to standard (non-predictive) runtime enforcement. All our results related to predictive RE of untimed properties are also formalized and proved in the Isabelle theorem prover. We also discuss how our predictive runtime enforcement framework can be extended to enforce timed properties.

The journal paper [24], done in collaboration with LaBRI Bordeaux and Inria Corse/LIG Grenoble, deals with runtime enforcement of untimed and timed properties with uncontrollable events. Runtime enforcement consists in defining and using mechanisms that modify the executions of a running system to ensure their correctness with respect to a desired property. We introduce a framework that takes as input any regular (timed) property described by a deterministic automaton over an alphabet of events, with some of these events being uncontrollable. An uncontrollable event cannot be delayed nor intercepted by an enforcement mechanism. Enforcement mechanisms should satisfy important properties, namely soundness, compliance, and optimality—meaning that enforcement mechanisms should output as soon as possible correct executions that are as close as possible to the input execution. We define the conditions for a property to be enforceable with uncontrollable events. Moreover, we synthesise sound, compliant, and optimal descriptions of runtime enforcement mechanisms at two levels of abstraction to facilitate their design and implementation.

#### 7.2.3.4. Control of logico-numerical systems.

In paper [32], we have targeted the problem of the safe control of reconfigurations in component-based software systems, where strategies of adaptation to variations in both their environment and internal resource demands need to be enforced. In this context, the computing system involves software components that are subject to control decisions. We have approached this problem under the angle of discrete-event systems (DES), involving properties on events observed during the execution (e.g., requests of computing tasks, work overload), and a state space representing different configurations such as activity or assemblies of components. We have considered in particular the potential of applying novel logico-numerical control techniques to extend the expressivity of control models and objectives, thereby extending the application of DES in componentbased software systems. We elaborate methodological guidelines for the application of logico-numerical control based on a case-study, and validate the result experimentally.

#### 7.2.4. Smart regulation for urban trains

Participants : Éric Fabre, Loïc Hélouët, Hervé Marchand, Karim Kecir

The regulation of subway lines consists in accomodating small random perturbations in transit times as well as more impacting incidents, by playing on continuous commands (transit times and dwell times) and by making more complex decisions (insertions or extractions of trains, changes of missions, overpassing, shorter returns, etc.) The objectives are multiple: ensuring the regularity and punctuality of trains, adapting to transportation demand, minimizing energy consumption, etc. We have developed an event-based control strategy that aims at equalizing headways on a line. This distributed control strategy is remarquably robust to perturbations and reactive enough to accomodate train insertions/extractions. We have integrated this control startegy to our SIMSTORS software. We have also developed another approach based on event graphs in order to optimally interleave trains at a junction. We started investigating new predictive control policies based of optimisation of criteria in forecast schedules [43].

In [47], we have extended a work started in 2016, that considers realizability of schedules by metro systems. Schedules are defined as high-level views of desired executions of systems, and represented as partial orders decorated with timing constraints. Train networks are modeled as stochastic time Petri nets (STPN) with an elementary (1-bounded) semantics. We have proposed a notion of time processes to give a partial-order semantics to STPNs. We then have considered Boolean realizability: a schedule S is realizable by a net Nif S embeds in a time process of N that satisfies all its constraints. However, with continuous time domains, the probability of a time process with exact dates is null. We thus consider probabilistic realizability up to  $\alpha$ time units, that holds if the probability that N realizes S with constraints enlarged by  $\alpha$  is strictly positive. Upon a sensible restriction guaranteeing time progress, Boolean and probabilistic realizability of a schedule can be checked on the finite set of symbolic prefixes extracted from a bounded unfolding of the net. We give a construction technique for these prefixes and show that they represent all time processes of a net occurring up to a given maximal date. We then show how to verify existence of an embedding and compute the probability of its realization. The technique has then been illustrated by a concrete example, namely deciding wheter a simple flip-flop shunting mechanism suffices to route trains in appropriate direction when delays can occur in trips or during stops at stations. We have also conducted a series of experiment [28] with the SIMSTORS tool to obtain statistics, and show feasibiliy of Key Performance Indicators (KPIs) evaluation with this formal model.

A second line of research relates to the development of new regulation strategies. New techniques were derived to equalize headways of trains along a line, and thus improve regularity and resilience to perturbations. A distributed control strategy was developed, easily implementable in existing rule engines. Simulations have proved the efficiency of this technique on orbital lines. We have also developed a global regulation approach based on timed event graphs. In this setting, control is event-based: a command is issued each time a train crosses a control point, but it takes into account information along the whole line and for a finite time horizon. This amounts to adapting the whole time-table for any new event in the system. This approach has been proved to perform well at junctions (on computer simulations), where randomly spaced trains arriving from two branches must be correctly interleaved at the junction of the two lines, while at the same time train intervals

must be equalized in all branches. We are now working on the combinatorial aspects of the question, in order to reduce energy consumption (by synchronizing arrivals and departures of trains), and in order to allow for insertions/extractions and reorderings of trains.

Several patents are in preparation for this activity.

#### 7.3. Management of Large Distributed Systems

#### 7.3.1. Analysis and synthesis of distributed systems

Participants : Éric Badouel, Thierry Jéron, Hervé Marchand, The Anh Pham

#### 7.3.1.1. Control of Distributed Systems.

In [40], we have extended our examination of decentralized discrete-event-system architectures that use exclusive or (XOR) as the fusion rule to reach control decisions. A characterization of XOR inference-observable languages has been provided. Additionally, XOR observability is defined for languages that are not inference-observable but are distributed-observable.

#### 7.3.1.2. Verification of distributed applications

In the context of IPL HAC-SPECIS, in collaboration with Martin Quinson (Myriads Inria project team) we are interested in the verification of real distributed applications.

In the conference paper [38] we explain the current status of the tool SimGridMC used for the verification of MPI applications. SimGridMC (also dubbed Mc SimGrid) is a stateful Model Checker for MPI applications. It is integrated to SimGrid, a framework mostly dedicated to predicting the performance of distributed applications. We describe the architecture of McSimGrid, and show how it copes with the state space explosion problem using Dynamic Partial Order Reduction and State Equality algorithms. As case studies we show how SimGrid can enforce safety and liveness properties for MPI applications, as well as global invariants over communication patterns.

#### 7.3.2. Analysis of parameterized systems

Participants : Nathalie Bertrand, Éric Fabre, Blaise Genest, Matthieu Pichené, Ocan Sankur

#### 7.3.2.1. Parameterized Verification of a time-synchronization protocol.

In [41], we consider distributed timed systems that implement leader-election protocols, which are at the heart of clock-synchronization protocols. We develop abstraction techniques for parameterized model checking of such protocols under arbitrary network topologies, where nodes have independently-evolving clocks. We apply our technique for model checking the root election part of the flooding time-synchronisation protocol (FTSP), and obtain improved results compared to previous work. We model-check the protocol for all topologies in which the distance to the node to be elected leader is bounded by a given parameter.

#### 7.3.2.2. Controlling population models.

In [33], we introduce a new setting where a population of agents, each modelled by a finite-state system, are controlled uniformly: the controller applies the same action to every agent. The framework is largely inspired by the control of a biological system, namely a population of yeasts, where the controller may only change the environment common to all cells. We study a synchronisation problem for such populations: no matter how individual agents react to the actions of the controller, the controller aims at driving all agents synchronously to a target state. The agents are naturally represented by a non-deterministic finite state automaton (NFA), the same for every agent, and the whole system is encoded as a 2-player game. The first player (Controller) chooses actions, and the second player (Agents) resolves non-determinism for each agent. The game with m agents is called the m-population game. This gives rise to a parameterized control problem (where control refers to 2-player games), namely the population control problem: can Controller control the m-population game for all  $m \in \mathbb{N}$ , whatever Agents does?

In this work, we prove that the population control problem is decidable, and it is an *EXPTIME*-complete problem. As far as we know, this is one of the first results on parameterized control. Our algorithm, not based on cut-off techniques, produces winning strategies which are symbolic, that is, they do not need to count precisely how the population is spread between states. We also show that if there is no winning strategy, then there is a population size M such that Controller wins the m-population game if, and only if,  $m \leq M$ . Surprisingly, M can be doubly-exponential in the number of states of the NFA, with tight upper and lower bounds.

7.3.2.3. Handling large biological systems.

This year, we propose to use approximated probabilistic distribution to handle large homogeneous populations of cells [39]. Beyond classical approximations, we propose to use the Chow-Liu tree representation, based on *non-disjoint* clusters of two variables. Our experiments show that our proposed approximation scheme is more accurate than existing ones to model probability distributions deriving from biopathways, while requiring a minimal complexity overhead.

To handle *dynamics* of a population of cells governed by biopathways, we develop *coarse-grained* abstractions of the biological pathways [21], and more precisely *Dynamic Bayesian Networks* (DBNs). We show that simulating a DBN is much faster than simulating the fine-grained model it abstracts, for comparable prediction performances.

We also explore the approximate inference problem of DBNs, that is, *computing* the probability distributions at every time point given the initial distribution at time 0. We evaluate several classical approximate inference algorithms for DBNs, and compare with a new method we propose, which consists in using the Chow-Liu tree approximation to represent distributions at each time step. It is very accurate, yet efficient according to experiments we report. We finally provide an error analysis of this approximate inference algorithm [39].

#### 7.4. Data-Driven Systems

#### 7.4.1. Incremental process discovery using Petri-net synthesis.

Participants : Éric Badouel

In [16], we present an incremental process discovery using Petri-net synthesis. Process discovery aims at constructing a model from a set of observations given by execution traces (a log). Petri nets are a preferred target model in that they produce a compact description of the system by exhibiting its concurrency. This article presents a process-discovery algorithm using Petri-net synthesis, based on the notion of region introduced by A. Ehrenfeucht and G. Rozenberg, and using techniques from linear algebra. The algorithm proceeds in three successive phases which make it possible to find a compromise between the ability to infer behaviours of the system from the set of observations while ensuring a parsimonious model, in terms of fitness, precision and simplicity. All used algorithms are incremental which means that one can modify the produced model when new observations are reported without reconstructing the model from scratch.

#### 7.4.2. An artifact model with imprecision and uncertainty

Participants : Éric Badouel, Loïc Hélouët

In the context of the HeadWork ANR project, we started investigating how complex workflows can be defined to handle uncertainty, and use joint knowledge of pools of user to build correct information. The solution proposed so far is a variant of business artifact managing fuzzy datasets. As there are several ways to reach an acceptable final and sufficiently precise dataset, we started investigating equivalence of complex workflows with partial information to allow refinement, enhance performance of data collection, with mastered precision loss.

## 8. Bilateral Contracts and Grants with Industry

#### 8.1. Bilateral Contracts with Industry

#### 8.1.1. ADR Softwarization of Everything

Joint Nokia-Inria research lab: Several researchers of SUMO are involved in the joint research lab of Nokia Bell Labs France and Inria, in a common research team called "Softwarization of Everything". The objective of this joint team is to design programming and management methods for software defined networks. Several other Inria teams take part to this group: Convecs, Diverse, Spades. Within this team, SUMO focuses on the management of reconfigurable systems, both at the edge (IoT based applications) and in the core (e.g. virtualized IMS systems). In particular, we focus on control and diagnosis issues for such systems.

#### 8.1.2. Alstom P22

Joint Alstom-Inria research lab: Several researchers of SUMO are involved in the joint research lab of Alstom and Inria, in a common research team called P22. On Alstom side, this joint research team involves researchers of the ATS division (Automatic Train Supervision). The objective of this joint team is to evaluate regulation policies of urban train systems, to assess their robustness to perturbations and failures, to design more efficient regulation policies and finally to provide decision support for human regulators. The project started in march 2014. A second phase of the project started in 2016, for a duration of three years. This covers in particular the CIFRE PhD of Karim Kecir.

## 9. Partnerships and Cooperations

#### 9.1. National Initiatives

## 9.1.1. ANR STOCH-MC: Model-Checking of Stochastic Systems using approximated algorithms (2014-2018)

- web site at http://perso.crans.org/~genest/stoch.html.
- Led by Blaise Genest (SUMO);
- Participants: Nathalie Bertrand, Blaise Genest, Éric Fabre, Matthieu Pichené;
- Partners: Inria Project Team CONTRAINTES (Rocquencourt), LaBRI (Bordeaux), and IRIF (Paris).

The aim of STOCH-MC is to perform model-checking of large stochastic systems, using controlled approximations. Two formalisms will be considered: Dynamic Bayesian Networks, which represent compactly large Markov Chains; and Markov Decision Processes, allowing non deterministic choices on top of probabilities.

#### 9.1.2. ANR HeadWork: Human-Centric Data-oriented WORKflows (2016-2020)

- web site at http://headwork.gforge.inria.fr/
- Led by David Gross-Amblard (Université Rennes 1);
- Participants : Loïc Hélouët, Éric Badouel;
- Partners: Inria Project-Teams Valda (Paris), DRUID (Rennes) SUMO (Rennes), LINKs (Lille), MNHN, Foule Factory.

The objective of this project is to develop techniques to facilite development, deployment, and monitoring of crowd-based participative applications. This requires handling complex workflows with multiple participants, incertainty in data collections, incentives, skills of contributors, ... To overcome these challenges, Headwork will define rich workflows with multiple participants, data and knowledge models to capture various kind of crowd applications with complex data acquisition tasks and human specificities. We will also address methods for deploying, verifying, optimizing, but also monitoring and adapting crowd-based workflow executions at run time.

#### 9.1.3. IPL HAC-SPECIS: High-performance Application and Computers, Studying PErformance and Correctness In Simulation (2016-2020)

- web site at http://hacspecis.gforge.inria.fr/
- Led by Arnaud Legrand (Inria Rhône-Alpes)
- Participants: Thierry Jéron, The Anh Pham.
- Partners: Inria project-teams Avalon (Lyon), POLARIS (Grenoble), HiePACS, STORM (Bordeaux), MExICo (Saclay), MYRIADS, SUMO (Rennes), VeriDis (Nancy).

The Inria Project Lab HAC-SPECIS (High-performance Application and Computers, Studying PErformance and Correctness In Simulation, 2016-2020: http://hacspecis.gforge.inria.fr/) is a transversal project internal to Inria. The goal of the HAC SPECIS project is to answer the methodological needs raised by the recent evolution of HPC architectures by allowing application and runtime developers to study such systems both from the correctness and performance point of view. Inside this project, we collaborate with Martin Quinson (Myriads team) on the dynamic formal verification of high performance runtimes and applications. The PhD of The Anh Pham is granted by this project.

This year we have been mainly intrested in dynamic partial-order-reduction methods that allow to reduce the explored state space, and a first prototype implementation of an existing method that combines DPOR with true-concurrency models.

#### 9.1.4. CNRS INS2I JCJC SensAs (2017)

- Led by Ocan Sankur (SUMO).
- Participants: Ocan Sankur
- Partners: Benjamin Monmege, Pierre-Alain Reynier (Université Aix-Marseille).

Model-checking allows one to analyse the reliability of critical systems. There is currently an ongoing effort to extend formal verification and synthesis techniques to check non-functional properties such as performance, energy consumption or robustness, that are particularly important for real-time systems. SensAS is a project whose objective is to develop techniques to analyse the sensitivity of such systems with formal tools. In this context, a nominal behaviour, described with a deterministic timed automaton, is submitted to nondeterministic or stochastic perturbations. We seek then to quantify the variability of perturbed behaviours, giving formal guarantees on the computed result.

#### 9.1.5. National informal collaborations

The team collaborates with the following researchers:

- Arnaud Sangnier (IRIF, UP7-Diderot) on the parameterized verification of probabilistic systems;
- François Laroussinie (IRIF, UP7-Diderot) on logics for multi-agent systems;
- Béatrice Bérard (LIP6) on problems of opacity and diagnosis, and on problems related to logics and partial orders for security;
- Serge Haddad (Inria team MExICo, LSV, ENS Paris-Saclay) on opacity and diagnosis;
- Patricia Bouyer (LSV, ENS Paris-Saclay) on the analysis of probabilistic timed systems and quantitative aspects of verification;
- Stefan Haar and Thomas Chatain (Inria team MExICo, LSV, ENS Paris-Saclay) on topics related to concurrency and time, and to modeling and verification of metro networks, multimodal systems and passenger flows;
- Éric Rutten and Gwenaël Delaval (Inria team Ctrl-A, LIG, Université Grenoble-Alpes) on the control of reconfigurable systems as well as making the link between Reax and Heptagon/BZR (http://bzr. inria.fr/);
- Didier Lime, Olivier H. Roux (LS2N Nantes) on topics related to stochastic and timed nets;
- Loïg Jezequel (LS2N Nantes) on topics related to stochastic and timed nets, and on distributed optimal planning;
- Yliès Falcone (CORSE LIG/Inria Grenoble) and Antoine Rollet (LaBRI Bordeaux) on the enforcement of timed properties;

## 9.2. International Initiatives

#### 9.2.1. Inria Associate Teams Not Involved in an Inria International Labs

#### 9.2.1.1. QuantProb

- Title: Quantitative analysis of non-standard properties in probabilistic models
- International Partner (Institution Laboratory Researcher):
  - Technical University of Dresde (Germany) Faculty of Computer Science Christel Baier
- Start year: 2016
- See also: http://www.irisa.fr/sumo/QuantProb/
- Quantitative information flow and fault diagnosis share two important characteristics: quantities (in the description of the system as well as in the properties of interest), and users partial knowledge. Yet, in spite of their similar nature, different formalisms have been proposed. Beyond these two motivating examples, defining a unified framework can be addressed by formal methods. Formal methods have proved to be effective to verify, diagnose, optimize and control qualitative properties of dynamic systems. However, they fall short of modelling and mastering quantitative features such as costs, energy, time, probabilities, and robustness, in a partial observation setting. This project proposal aims at developing theoretical foundations of formal methods for the quantitative analysis of partially observable systems.

#### 9.2.2. Inria International Partners

#### 9.2.2.1. Informal International Partners

The team collaborates with the following researchers:

- Jean-François Raskin, Gilles Geeraerts (Université Libre de Bruxelles, Belgium) on multiplayer game theory and synthesis;
- Thomas Brihaye (UMons, Belgium) on the verification of stochastic timed systems;
- Mickael Randour (UMons, Belgium) on quantitative games for synthesis;
- Kim G. Larsen (Aalborg University, Denmark) on quantitative timed games, and on topics related to urban train systems modeling;
- Josef Widder, Igor Konnov and Marijana Laźic (TU Wien, Austria) on the automated verification of randomized distributed algorithms.
- John Mullin (Polytechnique Montréal, Canada), on topics related to security and opacity;
- S. Akshay (IIT Bombay, India) on topics related to timed concurrent models;
- Andrea D'ariano (University Roma Tre, Italy), on topics related to train regulation;
- Stavros Tripakis, Srinivas Pinisetty (Aalto University, Finland) on runtime verification and enforcement.

#### 9.3. International Research Visitors

#### 9.3.1. Visits of International Scientists

- Laurie Ricker visited the SUMO team for 2 months in May-June 2017.
- 9.3.1.1. Internships
  - M2 Internship of Aina Toky Rasoamanana, Feb-July 2017, Nathalie Bertrand and Nicolas Markey
  - L3 Internship of Balasubramanian A.R., May-July 2017, Nathalie Bertrand and Nicolas Markey

#### 9.3.2. Visits to International Teams

9.3.2.1. Research Stays Abroad

- Éric Badouel made in September 2017 a one-month visit to Luca Bernardinello and Lucia Pomello from Milan University, and Carlo Ferigato from EJCR at Ispra. A work has been initiated on computer tools for the coordination of debates (from open citizen debates to parliamentary debates) and for managing the related documents (minutes, syntheses, ...) in an open data perspective.
- Engel Lefaucheux spent 6 weeks (May-June 2017) in Cagliari, working with Alessandro Giua and Carla Seatzu on the diagnosis of stochastic Petri nets.

## **10.** Dissemination

#### **10.1. Promoting Scientific Activities**

#### 10.1.1. Scientific Events Organisation

- 10.1.1.1. General Chair, Scientific Chair
  - Éric Badouel was the scientific chair of CRI 2017.
  - Hervé Marchand is member of the IFAC Technical Committees (TC 1.3 on Discrete Event and Hybrid Systems) since 2005. Hervé Marchand is member of the steering committee of MSR (modélisation de systèmes réactifs) since 2012 and became president of this steering in November 2017;
  - Nathalie Bertrand and Nicolas Markey are members of the steering committee of the Summer School MOVEP ("*Modelisation et Vérification des Processus Parallèles*").
- 10.1.1.2. Member of the Organizing Committees
  - Thierry Jéron is member of the steering committee of FMF 2017 (Formal Methods Forum), a forum gathering people from academia and industry and dedicated to the use of formal methods. It is held in Toulouse and, since this year, retransmitted in Grenoble, Saclay and Rennes. Two sessions took place in 2017, in January on the theme "Formal methods and cybersecurity" and in October about "Autonomous vehicles and formal methods".

#### 10.1.2. Scientific Events Selection

- 10.1.2.1. Chair of Conference Program Committees
  - Nathalie Bertrand was PC co-chair with Luca Bortolussi of QEST'2017, the 14th International Conference on Quantitative Evaluation of Systems, held in Berlin in September 2017 [45].
- 10.1.2.2. Member of the Conference Program Committees
  - Éric Badouel was member of the program committees of VECOS 2017, ATAED 2017 and CRI 2017.
  - Nathalie Bertrand was member of the PC of the following conferences: LICS 2017, FCT 2017, MSR 2017.
  - Blaise Genest was a PC member of ATVA 2017;
  - Thierry Jéron served on the Program Committees of the following international conferences: SAC-SVT 2017 and SAC-SVT 2018.
  - Hervé Marchand served on the Program Committees of MSR 2017, Wodes 2018, CDC 2017, CCTA 2017.
  - Nicolas Markey was a PC member of ATVA 2017, FORMATS 2017, SR 2017.
  - Ocan Sankur was a PC member of SR 2017 and SYNT 2017.

#### 10.1.3. Journal

#### 10.1.3.1. Member of the Editorial Boards

• Éric Badouel is co-editor-in-Chief of ARIMA Journal.

#### 10.1.4. Invited Talks

- Nathalie Bertrand was invited to give a talk at the workshop organized for the 20 years of LSV (ENS Cachan) in May 2017. She was also invited to talk at the workshop OPCT (Open Problems in Concurrency) in June 2017 at IST Vienna (Austria).
- Hervé Marchand gave an invited talk during the conference of MSR (November 2017) titled *Contribution to the Analysis of Discrete Event Systems* as well as during the workshop "30 years of the Ramadge-Wonham Theory of Supervisory Control: A Retrospective and Future Perspectives" at the CDC conference in December 2017 titled *Opacity and Supervisory Control.*
- Nicolas Markey gave an invited talk about *Temporal logics for multi-agent systems* at MFCS 2017, Aalborg (Denmark), in August 2017 [27].

#### 10.1.5. Leadership within the Scientific Community

• Since September 2017, Nathalie Bertrand is, together with Pierre-Alain Reynier, co-head of the *Groupe de Travail Verif* belonging to the *GDR Informatique Mathématique* (GDR-IM).

#### 10.1.6. Research Administration

- Éric Badouel is the co-director (with Moussa Lo, UGB, Saint-Louis du Sénégal) of LIRIMA, the Inria International Lab for Africa. He is scientific officer for the African and Middle-East region at Inria European and International Partnership Department and member of the executive board of GIS SARIMA.
- Nathalie Bertrand is elected member of the *Conseil National des Universités*, section 27 (computer science).
- Nathalie Bertrand, Loïc Hélouët and Ocan Sankur organize the weekly seminar 68NQRT at IRISA (40 talks each year).
- Éric Fabre is the co-director (with Olivier Audouin, Nokia) of the joint lab of Nokia Bell Labs France and Inria. The lab has been running for 9 years and started in Nov. 2017 its 3rd phase of joint research teams. A series of 6 new teams just started, for a duration of 4 years. They cover topics like network virtualization, network management, information theory, (distributed) machine learning, network security. SUMO is involved in the joint team "Softwarization of Everything".
- Loïc Hélouët is a representative of rank-B researchers in the *Comité de Centre* of Inria Rennes. He is also part of the bureau of the Comité de Centre. He leads the P22 projects with Alstom transports and is responsible for Workpackage 2 of the Headwork ANR project.
- Thierry Jéron is Member Committee Substitute for COST IC1402 ARVI (Runtime Verification beyond Monitoring). He is member of the IFIP Working Group 10.2 on Embedded Systems. He is member of the COS Prospective of Irisa Rennes and member of the *Comité de Centre* of Inria Rennes. Since 2016 he is *référent chercheur* for the Inria-Rennes research center.
- Hervé Marchand is chairman of the CUMI in Rennes and member of the ADT commission in Rennes.

#### **10.2. Teaching - Supervision - Juries**

#### 10.2.1. Teaching

- Licence: Nathalie Bertrand, Advanced Algorithms (ALGO2), 20h, L3, Univ Rennes 1, France;
- Licence: Loïc Hélouët, JAVA and algorithmics, L2, 40h, INSA de Rennes, France.
- Licence: Loïc Hélouët, practical studies (development of a small project), 8h, INSA de Rennes, France.
- Master : Nathalie Bertrand, Language Theory; Algorithms, 15h, Agrégation, ENS Rennes, France;
- Master: Éric Fabre, Models and Algorithms for Distributed Systems (MADS), 10h, M2, Univ Rennes 1, France;
- Master: Éric Fabre, Information Theory, 15h, M1, ENS Rennes, France.
- Master: Blaise Genest, Verification of Complex Systems (CSV), 10h, M2, Univ. Rennes 1, France;
- Master: Loïc Hélouët, Algorithms; complexity, 8h, Agrégation, ENS Rennes, France;
- Master: Loïc Hélouët, Nathalie Bertrand, Ocan Sankur, supervision of 3 students in M1 SIF (2017-2018).
- Master: Nicolas Markey, Verification of Complex Systems (CSV), 10h, M2, Univ Rennes 1, France;
- Master: Nicolas Markey, Algorithms for graphs, 3h, Agrégation, ENS Rennes, France;
- Master: Ocan Sankur, Lab sessions for the course on Foraml Analysis and Design (ACF), 22h, M1, Univ. Rennes 1, France.

#### 10.2.2. Supervision

10.2.2.1. Defences

- HdR: Hervé Marchand, *Contribution to the Analysis of Discrete Event Systems* [15], Univ. Rennes 1. The defence took place on 6 June 2017.
- PhD: Bruno Karelovic, Analyse quantitative des systèmes stochastiques Jeux de priorité et population de chaînes de Markov, Univ. Paris 7. The defence took place on 7 July 2017. Co-directed by Blaise Genest.

#### 10.2.2.2. PhD in progress

- Robert Nsaibirni, A Guarded Attribute Grammar Model for User centered Distibuted Collaborative Case Management: Case of the Disease Surveillance Process, co-advised by Éric Badouel and Maurice Tchuenté (University of Yaoundé).
- Engel Lefaucheux, *Controlling information in probabilistic systems*, started September 2015, Nathalie Bertrand and Serge Haddad
- The Anh Pham, *Dynamic Formal Verification of High Performance Runtimes and Applications*, started Nov. 2016, Thierry Jéron and Martin Quinson (Myriads, Inria Rennes).
- Karim Kecir, *Régulation et robustesse des systèmes ferroviaires urbains*, planned May 2018, Loïc Hélouët and Pierre Dersin (Alstom).
- Hugo Bazille, *Information flows in quantitative dynamic systems*, started oct. 2016, Blaise Genest and Éric Fabre.
- Erij Elmajed, *Diagnosis of reconfigurable systems*, started March 2017, Éric Fabre and Armen Aghasaryan (Nokia).
- Sihem Cherrared, *Diagnosis of multi-tenant programmable networks*, started Dec. 2016, Éric Fabre, Gregor Goessler (Inria, Spades) and Sofiane Imadali (Orange).
- Victor Roussanaly, *Efficient verification of timed systems*, started Sep. 2017, Nicolas Markey and Ocan Sankur.

#### 10.2.2.3. Master2 internship supervision

- Internship Aina Toky Rasoamanana, Feb-July 2017, Nathalie Bertrand and Nicolas Markey
- Internship Victor Roussanaly, Feb-June 2017, Nicolas Markey and Ocan Sankur

#### 10.2.2.4. Other internship supervision

- L3 Internship of Balasubramanian A.R., May-July 2017, Nathalie Bertrand and Nicolas Markey
- L3 Internship of Thomas Mari, Observation-based unfolding of Petri nets, (May-July 2017)
- L3 Internship of Romain Boitard, *Design of interfaces for railway systems*, (April-June 2017)

#### 10.2.3. Juries

10.2.3.1. Juries of PhD defences:

- Nicolas David, École Centrale Nantes, october 2017 : Nathalie Bertrand examiner;
- Thomas Geffroy, Univ. Bordeaux, december 2017: Nathalie Bertrand examiner;
- Ludovic Hofer, Univ. Bordeaux, november 2017: Blaise Genest examiner;
- Daniel Stan, École Normale Supérieure Paris-Saclay, march 2017: Nathalie Bertrand reviewer;
- Serge Tembo, Telecom Bretagne, January 2017: Éric Fabre examiner.

#### 10.2.3.2. Other juries

• Ocan Sankur was in the computer science entrance exam jury of École Normale Supérieures and École Polytechnique.

#### **10.3.** Popularization

• Éric Badouel gave a talk at TEDx Lorient on digital democracy (coordination of citizen debates).

# 11. Bibliography

## Major publications by the team in recent years

- [1] S. AKSHAY, B. GENEST, L. HÉLOUËT, S. YANG.Regular Set of Representatives for Time-Constrained MSC Graphs, in "Information Processing Letters", 2012, vol. 112, n<sup>o</sup> 14-15, p. 592-598, http://hal.inria.fr/hal-00879825.
- [2] E. BADOUEL, M. A. BEDNARCZYK, A. M. BORZYSZKOWSKI, B. CAILLAUD, P. DARONDEAU. Concurrent Secrets, in "Discrete Event Dynamic Systems", 2007, vol. 17, n<sup>o</sup> 4, p. 425-446.
- [3] E. BADOUEL, L. BERNARDINELLO, P. DARONDEAU. Petri Net Synthesis, Springer, 2015, http://dx.doi.org/ 10.1007/978-3-662-47967-4.
- [4] A. BENVENISTE, E. FABRE, S. HAAR, C. JARD.Diagnosis of Asynchronous Discrete Event Systems: A Net Unfolding Approach, in "IEEE Transactions on Automatic Control", November 2003, vol. 48, n<sup>o</sup> 5, p. 714-727, RNRT project MAGDA [DOI: 10.1109/TAC.2003.811249], http://hal.inria.fr/inria-00638224.
- [5] N. BERTRAND, B. GENEST, H. GIMBERT. Qualitative Determinacy and Decidability of Stochastic Games with Signals, in "Proceedings of the 24th Annual IEEE Symposium on Logic in Computer Science (LICS'09)", Los Angeles, USA, August 2009, http://hal.archives-ouvertes.fr/hal-00356566.
- [6] N. BERTRAND, T. JÉRON, A. STAINER, M. KRICHEN.Off-line Test Selection with Test Purposes for Non-Deterministic Timed Automata, in "Logical Methods in Computer Science", October 2012, vol. 8, n<sup>o</sup> 4:8, p. 1-33, http://hal.inria.fr/hal-00744074.
- [7] J. DUBREIL, P. DARONDEAU, H. MARCHAND. *Supervisory Control for Opacity*, in "IEEE Transactions on Automatic Control", May 2010, vol. 55, n<sup>o</sup> 5, p. 1089-1100 [DOI: 10.1109/TAC.2010.2042008].
- [8] E. FABRE, A. BENVENISTE. Partial Order Techniques for Distributed Discrete Event Systems: why you can't avoid using them, in "Journal of Discrete Events Dynamical Systems", 2007, vol. 17, n<sup>o</sup> 3, p. 357-403.

- [9] E. FABRE. Trellis Processes: A Compact Representation for Runs of Concurrent Systems, in "Journal of Discrete Event Dynamical Systems", 2007, vol. 17, n<sup>o</sup> 3, p. 267-306.
- [10] E. FABRE, L. JEZEQUEL. *Distributed optimal planning: an approach by weighted automata calculus*, in "Proceedings of the 48th IEEE Conference on Decision and Control (CDC'09)", 2009, p. 211-216.
- [11] B. GAUDIN, H. MARCHAND. An Efficient Modular Method for the Control of Concurrent Discrete Event Systems: A Language-Based Approach, in "Discrete Event Dynamic System", 2007, vol. 17, n<sup>o</sup> 2, p. 179-209.
- [12] T. GAZAGNAIRE, B. GENEST, L. HÉLOUËT, P. THIAGARAJAN, S. YANG. Causal Message Sequence Charts, in "Theoretical Computer Science", 2009, 38, EA DST, http://hal.inria.fr/inria-00429538.
- [13] C. JARD, T. JÉRON. TGV: theory, principles and algorithms, in "International Journal on Software Tools for Technology Transfer", 2005, vol. 7, n<sup>o</sup> 4, p. 297-315.
- [14] B. JEANNET, T. JÉRON, V. RUSU, E. ZINOVIEVA. Symbolic Test Selection Based on Approximate Analysis, in "Proceedings of the 11th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS'05)", Edinburgh, UK, 2005, http://hal.inria.fr/inria-00564617.

#### **Publications of the year**

#### **Doctoral Dissertations and Habilitation Theses**

[15] H. MARCHAND. Contribution to the Analysis of Discrete Event Systems, Université de Rennes 1, June 2017, Habilitation à diriger des recherches, https://hal.inria.fr/tel-01589972.

#### **Articles in International Peer-Reviewed Journal**

- [16] E. BADOUEL, U. SCHLACHTER. Incremental Process Discovery using Petri Net Synthesis, in "Fundamenta Informaticae", June 2017, vol. 154, n<sup>o</sup> 1-4, p. 1-13 [DOI: 10.3233/FI-2017-1548], https://hal.inria.fr/hal-01599760.
- [17] N. BERTRAND, B. GENEST, H. GIMBERT. Qualitative Determinacy and Decidability of Stochastic Games with Signals, in "Journal of the ACM (JACM)", October 2017, vol. 64, n<sup>o</sup> 5, p. 1 - 48 [DOI: 10.1145/3107926], https://hal.inria.fr/hal-01635127.
- [18] A. BOHY, V. BRUYÈRE, J.-F. RASKIN, N. BERTRAND.Symblicit algorithms for mean-payoff and shortest path in monotonic Markov decision processes, in "Acta Informatica", September 2017, vol. 54, n<sup>o</sup> 6, p. 545 -587 [DOI: 10.1007/s00236-016-0255-4], https://hal.inria.fr/hal-01635132.
- [19] B. BÉRARD, L. HÉLOUËT, J. MULLINS. Non-interference in partial order models, in "ACM Transactions on Embedded Computing Systems (TECS)", 2017, vol. 16, n<sup>o</sup> 2, 34 [DOI: 10.1145/2984639], https://hal. inria.fr/hal-01379451.
- [20] E. FABRE, L. HÉLOUËT, E. LEFAUCHEUX, H. MARCHAND. *Diagnosability of Repairable Faults*, in "Discrete Event Dynamic Systems", June 2017, https://hal.inria.fr/hal-01646911.

- [21] S. K. PALANIAPPAN, F. BERTAUX, M. PICHENÉ, E. FABRE, G. BATT, B. GENEST. Abstracting the dynamics of biological pathways using information theory: a case study of apoptosis pathway, in "Bioinformatics", February 2017, vol. 33, n<sup>o</sup> 13, p. 1980 - 1986, free access to online article: https://academic.oup.com/bioinformatics/article/33/13/1980/2996220/Abstractingthe-dynamics-of-biological-pathways?guestAccessKey=7083a393-ea30-4113-b5f9-816fb9b56287 [DOI: 10.1093/BIOINFORMATICS/BTX095], https://hal.inria.fr/hal-01547618.
- [22] S. PINISETTY, T. JÉRON, S. TRIPAKIS, Y. FALCONE, H. MARCHAND, V. PREOTEASA. Predictive Runtime Verification of Timed Properties, in "Journal of Systems and Software", October 2017, vol. 132, p. 353 - 365 [DOI: 10.1016/J.JSS.2017.06.060], https://hal.inria.fr/hal-01666995.
- [23] S. PINISETTY, V. PREOTEASA, S. TRIPAKIS, T. JÉRON, Y. FALCONE, H. MARCHAND. Predictive runtime enforcement, in "Formal Methods in System Design", August 2017, vol. 51, n<sup>o</sup> 1, p. 154 - 199 [DOI: 10.1007/s10703-017-0271-1], https://hal.inria.fr/hal-01647787.
- [24] M. RENARD, Y. FALCONE, A. ROLLET, T. JÉRON, H. MARCHAND. Optimal Enforcement of (Timed) Properties with Uncontrollable Events, in "Mathematical Structures in Computer Science", May 2017 [DOI: 10.1017/S0960129517000123], https://hal.archives-ouvertes.fr/hal-01262444.

#### **Invited Conferences**

- [25] S. AKSHAY, L. HÉLOUËT, R. PHAWADE. Combining Free choice and Time in Petri Nets, in "6th IFIP Working group on trends in Concurrency", Berlin, Germany, September 2017, Présentation dans un working group sans actes, https://hal.inria.fr/hal-01650751.
- [26] R. BRENGUIER, A. PAULY, J.-F. RASKIN, O. SANKUR. Admissibility in Games with Imperfect Information, in "CONCUR 2017 - 28th International Conference on Concurrency Theory", Berlin, Germany, LIPICs, September 2017, vol. 85 [DOI : 10.4230/LIPICS], https://hal.archives-ouvertes.fr/hal-01598152.
- [27] N. MARKEY.*Temporal Logics for Multi-Agent Systems*, in "MFCS 2017 42nd International Symposium on Mathematical Foundations of Computer Science", Aalborg, Denmark, August 2017 [DOI: 10.4230/LIPICS.MFCS.2017.84], https://hal.inria.fr/hal-01653855.

#### **International Conferences with Proceedings**

- [28] B. ADELINE, P. DERSIN, E. FABRE, L. HÉLOUËT, K. KECIR. An efficient evaluation scheme for KPIs in regulated urban train systems, in "RSSRail 2017 - International Conference on reliability, safety, and security of railway systems", Pistoia, Italy, A. FANTECHI, T. LECOMTE, A. ROMANOVSKY (editors), Lecture Notes in Computer Science, Springer, November 2017, n<sup>o</sup> 10598, https://hal.inria.fr/hal-01646919.
- [29] N. BASSET, G. GEERAERTS, J.-F. RASKIN, O. SANKUR. Admissibility in Concurrent Games, in "ICALP 2017 - 44th International Colloquium on Automata, Languages, and Programming", Warsaw, Poland, LIPICs, July 2017, vol. 80 [DOI: 10.4230/LIPICs.ICALP.2017.123], https://hal.archives-ouvertes.fr/hal-01598148.
- [30] N. BASSET, J.-F. RASKIN, O. SANKUR. Admissible Strategies in Timed Games, in "Models, Algorithms, Logics and Tools", aalborg, Denmark, Lecture Notes in Computer Science, July 2017, vol. 10460, https://hal. archives-ouvertes.fr/hal-01515874.

- [31] H. BAZILLE, E. FABRE, B. GENEST. Diagnosability Degree of Stochastic Discrete Event Systems, in "CDC 2017 - 56th IEEE Conference on Decision and Control", Melbourne, Australia, December 2017, p. 1-6, https:// hal.inria.fr/hal-01651232.
- [32] N. BERTHIER, F. ALVARES, H. MARCHAND, G. DELAVAL, E. RUTTEN.Logico-numerical Control for Software Components Reconfiguration, in "CCTA 2017 - IEEE Conference on Control Technology and Applications", Mauna Lani, HI, United States, IEEE, August 2017, p. 1599 - 1606 [DOI: 10.1109/CCTA.2017.8062685], https://hal.inria.fr/hal-01644754.
- [33] N. BERTRAND, M. DEWASKAR, B. GENEST, H. GIMBERT. Controlling a Population, in "CONCUR 2017 -28th International Conference on Concurrency Theory", Berlin, Germany, September 2017, p. 1-23, https:// hal.archives-ouvertes.fr/hal-01625661.
- [34] P. BOUYER, P. HOFMAN, N. MARKEY, M. RANDOUR, M. ZIMMERMANN.*Bounding Average-Energy Games*, in "FoSSaCS'17", Uppsala, Sweden, Proceedings of the 20th International Conference on Foundations of Software Science and Computation Structure (FoSSaCS'17), Springer, April 2017, vol. 10203, p. 179-195 [DOI: 10.1007/978-3-662-54458-7\_11], https://hal.archives-ouvertes.fr/hal-01566431.
- [35] P. BOUYER, S. JAZIRI, N. MARKEY. On the determinization of timed systems, in "FORMATS'17", Berlin, Germany, Proceedings of the 15th International Conferences on Formal Modelling and Analysis of Timed Systems (FORMATS'17), September 2017, vol. 10419, p. 25-41 [DOI : 10.1007/978-3-319-65765-3 2], https://hal.archives-ouvertes.fr/hal-01566436.
- [36] B. BÉRARD, S. HADDAD, E. LEFAUCHEUX. Probabilistic Disclosure: Maximisation vs. Minimisation, in "FSTTCS 2017", Kanpur, India, December 2017 [DOI: 10.4230/LIPICS.FSTTXS.2017], https://hal.inria. fr/hal-01618955.
- [37] L. HÉLOUËT, H. MARCHAND. On the cost of diagnosis with disambiguation, in "QEST 2017", Berlin, France, 14th International Conference on Quantitative Evaluation of SysTems (QEST 2017), September 2017, https:// hal.inria.fr/hal-01537796.
- [38] T. A. PHAM, T. JÉRON, M. QUINSON. Verifying MPI Applications with SimGridMC, in "Correctness 2017 First International Workshop on Software Correctness for HPC Applications", Denver, United States, November 2017 [DOI: 10.1145/3145344.3145345], https://hal.inria.fr/hal-01632421.
- [39] M. PICHENÉ, S. K. PALANIAPPAN, E. FABRE, B. GENEST.*Non-Disjoint Clustered Representation for Distributions over a Population of Cells (poster)*, in "CMSB 2017", Darmstadt, Germany, CMSB 2017 15th International Conference on Computational Methods in Systems Biology, Springer, 2017, vol. LNCS/LNBI, n<sup>o</sup> 10545, p. 324-326, https://hal.archives-ouvertes.fr/hal-01625665.
- [40] L. RICKER, T. F. LIDBETTER, H. MARCHAND.Inferencing and beyond: further adventures with parity-based architectures for decentralized discrete-event systems, in "20th World Congress The International Federation of Automatic Control", Toulouse, France, July 2017, 6, https://hal.inria.fr/hal-01590438.
- [41] O. SANKUR, J.-P. TALPIN. An Abstraction Technique For Parameterized Model Checking of Leader Election Protocols: Application to FTSP, in "23rd International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)", Uppsala, Sweden, Lecture Notes in Computer Science, April 2017, vol. 10206, https://hal.archives-ouvertes.fr/hal-01431472.

#### National Conferences with Proceeding

#### [42] Best Paper

N. BERTRAND, S. HADDAD, E. LEFAUCHEUX.*Diagnostic et contrôle de la dégradation des systèmes probabilistes*, in "MSR 2017 - Modélisation des Systèmes Réactifs", Marseille, France, November 2017, https://hal.inria.fr/hal-01618922.

#### **Conferences without Proceedings**

[43] K. KECIR, L. HELOUET, P. DERSIN, B. ADELINE, A. D'ARIANO. From Reactive to Predictive Regulation in Metros, in "ECSO 2017 : 2nd European Conference on Stochastic Optimization", Rome, Italy, September 2017, https://hal.inria.fr/hal-01646916.

#### Scientific Books (or Scientific Book chapters)

[44] P. BOUYER, F. LAROUSSINIE, N. MARKEY, J. OUAKNINE, J. WORRELL.*Timed temporal logics*, in "Models, Algorithms, Logics and Tools: Essays Dedicated to Kim Guldstrand Larsen on the Occasion of His 60th Birthday", Lecture Notes in Computer Science, Springer, August 2017, vol. 10460, p. 211-230 [DOI: 10.1007/978-3-319-65764-6\_11], https://hal.archives-ouvertes.fr/hal-01566439.

#### **Books or Proceedings Editing**

[45] N. BERTRAND, L. BORTOLUSSI (editors). Proceedings of the 14th International Conference on Quantitative Evaluation of Systems (QEST 2017), Berlin, Germany, September 5-7, 2017, 2017, https://hal.inria.fr/hal-01635134.

#### **Research Reports**

- [46] S. AKSHAY, L. HÉLOUËT, R. PHAWADE. Combining Free choice and Time in Petri Nets, Inria Rennes Bretagne Atlantique, July 2017, https://hal.inria.fr/hal-01646913.
- [47] L. HÉLOUËT, K. KECIR. *Realizability of Schedules by Stochastic Time Petri Nets with Blocking Semantics:* (*Extended Version*), Inria Rennes Bretagne Atlantique, October 2017, https://hal.inria.fr/hal-01646920.

#### **Other Publications**

- [48] N. BERTRAND, M. DEWASKAR, B. GENEST, H. GIMBERT. *Controlling a Population*, July 2017, https://arxiv. org/abs/1707.02058 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01558029.
- [49] P. GARDY, P. BOUYER, N. MARKEY. *Dependences in Strategy Logic*, October 2017, https://arxiv.org/abs/ 1708.05849 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01609523.
- [50] V. JUGÉ, P. BOUYER-DECITRE, N. MARKEY. Courcelle's Theorem Made Dynamic, October 2017, https:// arxiv.org/abs/1702.05183 - 14 pages, 4 figures. arXiv admin note: text overlap with arXiv:1610.00571, https:// hal.archives-ouvertes.fr/hal-01615275.

#### **References in notes**

[51] E. FABRE, L. HÉLOUËT, E. LEFAUCHEUX, H. MARCHAND.*Diagnosability of Repairable Faults*, in "13th International Workshop on Discrete Event Systems (WODES'16)", Xi'an, China, 2016, p. 256–262.

[52] F. MOGAVERO, A. MURANO, G. PERELLI, M. Y. VARDI.*Reasoning About Strategies: On the Model-Checking Problem*, in "ACM Transactions on Computational Logic", August 2014, vol. 15, n<sup>o</sup> 4, p. 34:1-34:47, http://dx.doi.org/10.1145/2631917.

# **Team TACOMA**

# **TAngible COMputing Architectures**

Inria teams are typically groups of researchers working on the definition of a common project, and objectives, with the goal to arrive at the creation of a project-team. Such project-teams may include other partners (universities or research institutions).

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Distributed programming and Software engineering

## **Table of contents**

1.	Personnel						
2.	• Overall Objectives						
3.	Research Program						
	3.1. Collecting pertinent information						
	3.2. Building relevant abstraction for new interactions	915					
	3.3. Acting on the environment	915					
4.	Application Domains						
	4.1. Pervasive applications in Smart Building	916					
	4.2. Metamorphic House	916					
	4.3. Automation in Smart City	917					
	4.4. Pervasive applications in uncontrolled environmements	917					
5.	New Software and Platforms	918					
	5.1. THEGAME	918					
	5.2. Platform Pervasive_RFID	918					
	5.3. Metamorphic Housing platform and Software - On-demand room	919					
	5.4. ISO/IEC 15118-2 Open source Implementation	920					
6.	New Results						
	6.1. Smart City and ITS	920					
	6.2. Convergence middleware for pervasive data	921					
	6.3. Modeling activities and forecasting energy consumption and production to promote the u	se of					
	self-produced electricity from renewable sources	922					
	6.4. Sharing knowledge and access-control	923					
7.	Bilateral Contracts and Grants with Industry	924					
8.	Partnerships and Cooperations	924					
	8.1. Regional Initiatives	924					
	8.2. European Initiatives	925					
9.	Dissemination	926					
	9.1. Promoting Scientific Activities	926					
	9.1.1. Scientific Events Organisation	926					
	9.1.2. Scientific Events Selection	926					
	9.1.2.1. Reviewer	926					
	9.1.2.2. Member of the Conference Program Committees	926					
	9.1.3. Invited Talks	926					
	9.1.4. Leadership within the Scientific Community	927					
	9.1.5. Scientific Expertise	927					
	9.1.6. Research Administration	927					
	9.2. Teaching - Supervision - Juries	927					
	9.2.1. Teaching	927					
	9.2.2. Supervision	927					
	9.2.3. Juries	927					
	9.3. Popularization	928					
10.	Bibliography	928					

### **Team TACOMA**

*Creation of the Team: 2014 January 01, end of the Team: 2017 December 31* **Keywords:** 

#### **Computer Science and Digital Science:**

- A1.2. Networks
- A1.2.5. Internet of things
- A1.2.6. Sensor networks
- A1.2.7. Cyber-physical systems
- A1.3. Distributed Systems
- A1.4. Ubiquitous Systems
- A2.3. Embedded and cyber-physical systems
- A2.3.2. Cyber-physical systems
- A2.5. Software engineering
- A2.5.1. Software Architecture & Design
- A2.5.3. Empirical Software Engineering
- A2.6. Infrastructure software
- A2.6.1. Operating systems
- A2.6.2. Middleware
- A4.8. Privacy-enhancing technologies
- A5.11. Smart spaces
- A5.11.1. Human activity analysis and recognition
- A5.11.2. Home/building control and interaction

#### **Other Research Topics and Application Domains:**

- B3.1. Sustainable development
- B3.1.1. Resource management
- B4.4. Energy delivery
- B4.4.1. Smart grids
- B4.5.2. Embedded sensors consumption
- B6.1. Software industry
- B6.1.1. Software engineering
- B6.1.2. Software evolution, maintenance
- B6.2.2. Radio technology
- B6.3.3. Network Management
- B6.4. Internet of things
- B6.6. Embedded systems
- B7.2. Smart travel
- B7.2.1. Smart vehicles
- B7.2.2. Smart road
- B8.1. Smart building/home
- B8.1.1. Energy for smart buildings
- B8.1.2. Sensor networks for smart buildings

B8.2. - Connected cityB8.5.2. - Crowd sourcingB8.5.3. - Collaborative economyB9.8. - PrivacyB9.10. - Ethics

# 1. Personnel

#### **Research Scientist**

Paul Couderc [Inria, Researcher]

#### **Faculty Members**

Jean-Marie Bonnin [Team leader, IMT Atlantique (Cesson-Sévigné), Professor, HDR] Yoann Maurel [Univ. Rennes I, Associate Professor] Frédéric Weis [Univ. Rennes I, Associate Professor, HDR]

#### **Technical Staff**

Michele Dominici [Univ. Rennes I, until Mar 2017] Xavier Gilles [IMT Atlantique (Cesson-Sévigné), from Mar 2017 until Nov 2017] Guillaume Habault [IMT Atlantique (Cesson-Sévigné), from Mar 2017 until Aug 2017] Guillaume Le Gall [IMT Atlantique (Cesson-Sévigné), until Nov 2017]

#### PhD Students

Christophe Couturier [IMT Atlantique (Cesson-Sévigné), PhD student] Adrien Capaine [OyaLight, PhD Student, granted by CIFRE] Indra Ngurah [IMT Atlantique (Cesson-Sévigné)] Alexandre Rio [OKWIND, granted by CIFRE] Rodrigo Silva [IMT Atlantique (Cesson-Sévigné)] Liouane Zaineb [Co-tutelle Univ. Rennes 1 with Univ. Monastir, until Oct 2017]

#### **Post-Doctoral Fellow**

Djibrilla Amadou Kountche [IMT Atlantique (Cesson-Sévigné), until Sep 2017]

#### Administrative Assistant

Fabienne Cuyollaa [Inria]

# 2. Overall Objectives

#### 2.1. Presentation

The technologies necessary for the development of pervasive applications are now widely available and accessible for many uses: short/long-range and low energy communications, a broad variety of visible (smart objects) or invisible (sensors and actuators) objects, as well as the democratization of the Internet of Things (IoT). Large areas of our living spaces are now instrumented. The concept of Smart Spaces is about to emerge, based upon both massive and apposite interactions between individuals and their everyday working and living environments: residential housing, public buildings, transportation, etc. The possibilities of new applications are boundless. Many scenarios have been studied in laboratories for many years and, today, a real ability to adapt the environment to the behaviors and needs of users can be demonstrated. However mainstream pervasive applications are barely existent, at the notable exception of the ubiquitous GPS-based navigators. The opportunity of using vast amount of data collected from the physical environments for **several application domains** is still largely untapped. The applications that interact with users and act according to their environment with a large autonomy are still very specialized. They can only be used in the environment they had especially been developed for (for example "classical" home automation tasks: comfort,

entertainment, surveillance). They are difficult to adapt to increasingly complex situations, even though the environments in which they evolve are more open, or change over time (new sensors added, failures, mobility etc.).

Developing applications and services that are ready to deploy and evolve in different environments should involve significant cost reduction. Unfortunately, designing, testing and ensuring the maintenance as well as the evolution of a pervasive application remains very complex. In our view, the lack of resources by which properties of the real environment are made available to application developers is a major concern. Building a pervasive application involves implementing one or more logical control loops which include four stages (see figure 1-a): (1) data collection in the real environment, (2) the (re)construction of information that is meaningful for the application and (3) for decision making, and finally, (4) action within the environment. While many decision-algorithms have been proposed, the **collection** and **construction** of a reliable and relevant perception of the environment and, in return, **action** mechanisms within the environment still pose major challenges that the TACOMA project is prepared to deal with.

Most current solutions are based on a massive collection of raw data from the environment, stored on remote servers. Figure 1-a illustrates this type of approach. Exposure of raw sensor values to the decision-making process does not allow to build relevant contexts that a pervasive application actually needs in order to shrewdly act/react to changes in the environment. So, the following is left up to the developer:

- To characterize more finely raw data beyond its simple value, for example, the acquisition date, the nature of network links crossed to access the sensor, the durability and accuracy of value reading, etc.
- To exploit this raw data to calculate a relevant abstraction for the application, such as, whether the room is occupied, or whether two objects are in the same physical vicinity.
- To modify the environment when possible.

Traditional software architectures isolate the developer from the real environment that he oftentimes has to depict according to complex, heavy and expensive processes. However, objects and infrastructure integrated into user environments could provide a more suitable support to pervasive applications: description of the actual system's state can be richer, more accurate, and, meanwhile, easier to handle; the applications' structure can be distributed by being built directly into the environment, facilitating scalability and resilience by the processing autonomy; finally, moving processing closer to the edge of the network avoids major problems of data sovereignty and privacy encountered in infrastructures very dependent on the cloud. We strongly believe in the advantages of specific approaches to the fields of **edge computing** and **fog computing**, which will reveal themselves with the development of Smart Spaces and an expansive growth of the number of connected objects. Indeed, ensuring the availability and reliability of systems that remain frugal in terms of resources will become in the end a major challenge to be faced in order to allow proximity between processing and end-users. Figure 1-b displays the principle of "using data at the best place for processing". Fine decisions can be made closer to the objects producing and acting on the data, local data characterization and local processing de-emphasize the computing and storage resources of the cloud (which can be used for example to store selected/transformed data for global historical analysis or optimization).

TACOMA aims at developing a comprehensive set of new **interaction models** and **system architectures** to considerably help pervasive application designers in the development phase with the side effect to ease the life cycle management. We follow two main principles:

- Leveraging local properties and direct interactions between objects, we would be able to enrich and to manage locally data produced in the environment. The application would then be able to build their knowledge about their environment (perception) in order to adjust their behavior (eg. level of automation) to the actual situation.
- Pervasive applications should be able to describe requirements they have on the quality of their environment perception. We would be able to achieve the minimum quality level adapting the diversity of the sources (data fusion/aggregation), the network mechanisms used to collect the data (network/link level) and the production of the raw data (sensors).



Figure 1. Adaptation processes in pervasive environments

# 3. Research Program

#### 3.1. Collecting pertinent information

In our model, applications adapt their behavior (for instance, the level of automation) to the quality of their perception of the environment. This is important to alleviate the development constraint we usually have on automated system. We "just" have to be sure a given process will always operate at the right automation level given the precision, the completeness or the confidence it has on its own perception. For instance, a car passing through a cross would choose its speed depending on the confidence it has gained during perception data gathering. When it has not enough information or when it could not trust it, it should reduce the automation level, therefore the speed, to only rely on its own sensors. Such adaptation capability shift requirements from the design and deployment (availability, robustness, accuracy, etc.) to the **assessment of the environment perception** we aim to facilitate in this first research axis.

*Data characterization.* The quality (freshness, accuracy, confidence, reliability, confidentiality, etc.) of the data are of crucial importance to assess the quality of the perception and therefore to ensure proper behavior. The way data is produced, consolidated, and aggregated while flowing to the consumer has an impact on its quality. Moreover part of these quality attributes requires to gather information at several communication layers from various entities. For this purpose, we want to design **lightweight cross-layer interactions** to collect relevant data. As a "frugality" principle should guide our approach, it is not appropriate to build all attributes we can imagine. It is therefore necessary to identify attributes relevant to the application and to have mechanisms to activate/deactivate at run-time the process to collect them.

*Data fusion.* Raw data should be directly used only to determine low-level abstraction. Further help in abstracting from low-level details can be provided by **data fusion** mechanisms. A good (re)construction of a meaningful information for the application reduces the complexity of the pervasive applications and helps

the developers to concentrate on the application logic rather on the management of raw data. Moreover, the reactivity required in pervasive systems and the aggregation of large amounts of data (and its processing) are antagonists. We study **software services that can be deployed closer to the edge of the network**. The exploration of data fusion technics will be guided by different criteria: relevance of abstractions produced for pervasive applications, anonymization of exploited raw data, processing time, etc.

Assessing the correctness of the behavior. To ease the design of new applications and to align the development of new products with the ever faster standard developments, continuous integration could be used in parallel with continuous conformance and interoperability testing. We already participate in the design of new shared platforms that aims at facilitating this providing remote testing tools. Unfortunately, it is not possible to be sure that all potential peers in the surrounding have a conform behavior. Moreover, upon failure or security breach, a piece of equipment could stop to operate properly and lead to global mis-behavior. We want to propose conceptual tools for **testing at runtime devices in the environment**. The result of such conformance or interoperability tests could be stored safely in the environment by authoritative testing entity. Then application could interact with the device with a higher confidence. The confidence level of a device could be part of the quality attribute of the information it contributed to generate. The same set of tools could be used to identify misbehaving device for maintenance purpose or to trigger further testing.

#### 3.2. Building relevant abstraction for new interactions

The pervasive applications are often designed in an ad hoc manner depending on the targeted application area. Ressources (sensors / actuators, connected objets etc.) are often used in silos which complexify the implementation of rich pervasive computing scenarios. In the second research axis, we want to get away from technical aspects identifying **common and reusable system mechanisms** that could be used in various applications.

*Tagging the environment*. Information relative to environment could be stored by the application itself, but it could be complex to manage for mobile application since it could cross a large number of places with various features. Moreover the developer has to build its own representation of information especially when he wants to share information with other instances of the same application or with other applications. A promising approach is to store and to maintain this information associated to an object or to a place, in the environment itself. The infrastructure should provide services to application developers: add/retrieve information in the environment, share information and control who can access it, add computed properties to object for further usage. We want to study an **extensible model to describe and augment the environment**. Beyond a simple distributed storage, we have in mind a new kind of interaction between pervasive applications and changing environment and between applications themselves.

Taking advantages of the spatial and temporal relationships. To understand the world they have to interact with, pervasive applications often have to (re)built a model of it from the exchange they have with others or from their own observations. A part of the programmer's task consists in building a model of the spatial layout of the objects in the surrounding. The term *layout* can be understood in several ways: the co-location of multiple objects in the same vicinity, the physical arrangement of two objects relative to each other, or even the crossing of an object of a physical area to another, etc. Determining remotely these spatial properties (see figure 1-a) is difficult without exchanging a lot of information. Properties related to the spatial layout are far easier to characterize locally. They could be abstracted from interaction pattern without any complex virtual representation of the environment (see figure 1-b). We want to be able to rely on this type of spatial layout in a pervasive environment. In the prior years, the members of TACOMA already worked on **models for processing object interactions** in the physical world to automatically trigger processing. This was the case in particular of the spatial programming principle: physical space is treated as a tuple-space in which objects are automatically synchronized according to their spatial arrangement. We want to follow this approach by considering **richer and more expressive programming models.** 

#### **3.3.** Acting on the environment

The conceptual tools we aim to study must be *frugal*: they use as less as possible resources, while having the possibility to use much more when it is required. Data needed by an application are not made available for "free"; for example, it costs energy to measure a characteristic of the environment, or to transmit it. So this "design frugality" requires **a fine-grained control** on how data is actually collected from the environment. The third research axis aims at designing solutions that give this control to application developers by **acting on the environment**.

Acting on the data collection. We want to be able to identify which information are reality needed during the perception elaboration process. If a piece of data is missing to build a given information with the appropriate quality level, the data collection mechanism should find relevant information in the environment or modify the way it aggregate it. These could lead to a modification of the behavior of the network layer and the path the piece of data use in the aggregation process.

Acting on object interactions. Object in the environment could adapt their behavior in a way that strongly depend on the object itself and that is difficult to generalize. Beyond the specific behaviors of actuators triggered through specialized or standard interfaces, the production of information required by an application could necessitate an adaptation at the object level (eg. calibration, sampling). The environment should then be able to initiate such adaption transparently to the application, which may not know all objects it passes by.

Adapting object behaviors. The radio communication layers become more flexible and able to adapt the way they use energy to what is really required for a given transmission. We already study how beamforming technics could be used to adapt multicast strategy for video services. We want to show how playing with these new parameters of transmissions (eg. beamforming, power, ...) allows to control spatial relationships objects could have. There is a tradeoff to find between the capacity of the medium, the electromagnetic pollution and the reactivity of the environment. We plan to expend our previous on interface selection and more generally on what we call **opportunistic networking**.

# 4. Application Domains

#### 4.1. Pervasive applications in Smart Building

A Smart Building is a living space equipped with information-and-communication-technology (ICT) devices conceived to collaborate in order to anticipate and respond to the needs of the occupants, working to promote their comfort, convenience, security and entertainment while preserving their natural interaction with the environment.

The idea of using the Pervasive Computing paradigm in the Smart Building domain is not new. However, the state-of-the-art solutions only partially adhere to its principles. Often the adopted approach consists in a heavy deployment of sensor nodes, which continuously send a lot of data to a central elaboration unit, in charge of the difficult task of extrapolating meaningful information using complex techniques. This is a *logical approach*. TACOMA proposed instead the adoption of a *physical approach*, in which the information is spread in the environment, carried by the entities themselves, and the elaboration is directly executed by these entities "inside" the physical space. This allows performing meaningful exchanges of data that will thereafter need a less complicated processing compared to the current solutions. The result is a smart environment that can, in an easier and better way, integrate the context in its functioning and thus seamlessly deliver more useful and effective user services. Our contribution aims at implementing the physical approach in a smarter environment, showing a solution for improving both comfort and energy savings.

#### 4.2. Metamorphic House

The motivation for metamorphic houses is that many countries, including France, are going through sociodemographic evolutions, like growth of life expectancy and consequent increase in the number of elderly people, urbanization and resource scarcity. Households experience financial restrictions, while housing costs increase with the raise of real estate and energy prices [6]. Important questions arise concerning the future of housing policies and ways of living. We observe novel initiatives like participative housing and developing behaviors, including house-sharing, teleworking and longer stay of children in parents' homes.

To tackle the challenges raised by these emerging phenomena, future homes will have to be modular, upgradeable, comfortable, sparing of resources. They should be integrated in the urban context and exchange information with other homes, contribute to reducing the distances to be covered daily and respect the characteristics of the territory where they are located.

To reach these goals, metamorphic domestic environments will modify their shape and behavior to support activities and changes in life cycle of occupants, increase comfort and optimize the use of resources. Thanks to Information and Communication Technologies (ICT) and adaptive building elements, the same physical spaces will be transformed for different uses, giving inhabitants the illusion of living in bigger, more adapted and more comfortable places.

#### 4.3. Automation in Smart City

The domain of Smart Cities is still young but it is already a huge market which attract number of companies and researchers. It is also multi-fold as the words "smart city" gather multiple meanings. Among them one of the main responsibilities of a city, is to organize the transportation of goods and people. In intelligent transportation systems (ITS), ICT technologies have been involved to improve planification and more generally efficiency of journeys within the city. We are interested in the next step where efficiency would be improved locally relying on local interactions between vehicles, infrastructure and people (smartphones).

For the future autonomous vehicle are now in the spotlight, since a lot of works has been done in recent years in automotive industry as well as in academic research centers. Such unmanned vehicle could strongly impact the organisation of the transportation in our cities. However, due to the lack of a definition of what is an "autonomous" vehicle it remains still difficult to see how these vehicles will interact with their environment (eg. road, smart city, houses, grid, etc.). From augmented perception to fully cooperative automated vehicle, the autonomy covers various realities in terms of interaction the vehicle relies on. The extended perception relies on communication between the vehicle and surrounding roadside equipments. This help the driving system to build and maintain an accurate view of the environment. But at this first stage the vehicle only uses its own perception to make its decisions. At a second stage, it will take advantages of local interaction with other vehicles through car-to-car communications to elaborate a better view of its environment. Such "cooperative autonomy" does not try to reproduce the human behavior anymore, it strongly relies on communication between vehicles and/or with the infrastructure to make decision and to acquire information on the environment. Part of the decision could be centralized (almost everything for an automatic metro) or coordinated by a roadside component. The decision making could even be fully distributed but this put high constraints on the communications. Automated vehicles are just an exemple of smart city automated processes that will have to share information within the surrounding to make their decisions.

#### 4.4. Pervasive applications in uncontrolled environnements

Some limitations of existing RFID technology become challenging: unlike standard RFID application scenarios, pervasive computing often involves uncontrolled environment for RFID, where tags and reader have to operate in much more difficult situations that those usually encountered or expected for classical RFID systems.

RFID technology is to avoid missing tags when reading multiple objects, as reading reliability is affected by various effects such shadowing or wave power absorption by some materials. The usual applications of RFID operate in a controlled environment in order to reduce the risk of missing tags while scanning objects.

In pervasive computing applications, a controlled reading environment is extremely difficult to achieve, as one of the principle is to enhance existing processes "in situ", unlike the controlled conditions that can be found in industrial processes. Consider for example a logistic application, where RFID tags could be used on items inside a package in order to check for its integrity along the shipping process. Tags would likely be placed randomly on items inside the package, and reading conditions would be variable depending on where the package is checked.

RFID operation in uncontrolled environments is challenging because RFID performance is affected by multiple parameters, in particular:

- Objects materials (on which tags are attached to),
- Materials in the surrounding environment,
- RFID frequency spectrum,
- Antenna nature and placement with respect to the tags.

In controlled environment, the difficulty to read tags can be limited by using the appropriate parameters to maximize the RFID performance for the application. But in many cases, it is needed to read large number of objects of various nature, arranged randomly in a given area or container. **Most pervasive computing applications fall in this context**.

# 5. New Software and Platforms

#### **5.1. THEGAME**

SCIENTIFIC DESCRIPTION: Context-aware applications have to sense the environment in order to adapt themselves and provide with contextual services. This is the case of Smart Homes equipped with sensors and augmented appliances. However, sensors can be numerous, heterogeneous and unreliable. Thus the data fusion is complex and requires a solid theory to handle those problems. The aim of the data fusion, in our case, is to compute small pieces of context we call context attributes. Those context attributes are diverse and could be for example the presence in a room, the number of people in a room or even that someone may be sleeping in a room. For this purpose, we developed an implementation of the belief functions theory (BFT). THE GAME (THeory of Evidence in a lanGuage Adapted for Many Embedded systems) is made of a set of C-Libraries. It provides the basics of belief functions theory, computations are optimized for an embedded environment (binary representation of sets, conditional compilation and diverse algorithmic optimizations).

THE GAME is published under apache licence (https://github.com/bpietropaoli/THEGAME/). It is maintained and experimented by Aurélien Richez within a sensor network platform developed by TACOMA since June 2013.

FUNCTIONAL DESCRIPTION: THEGAME is a set of software services for detecting different types of situation in a building (presence in a room, activity level, etc.) based on a set of raw data sourced from all sorts of sensors. Written in C or Java, it can be integrated in an embedded computer: tablet, smartphone, box, etc., and can be connected to different sensor networks. It can be used to implement context-aware services: for example, to alert the user if s/he forgets to close a window when leaving the building, or to turn off the heating in an empty room, etc.

- Participants: Aurélien Richez and Bastien Pietropaoli
- Contact: Frédéric Weis
- URL: https://github.com/bpietropaoli/THEGAME/

#### 5.2. Platform Pervasive\_RFID

**KEYWORDS:** Composite objects - RFID

- Participants: Paul Couderc and Anthony Blair (Univ. Rennes 1)
- Partner: Univ. Rennes 1 (IETR)
- Contact: Paul Couderc

#### SCIENTIFIC DESCRIPTION

The RFID experiment testbed has been designed and deployed in collaboration with IETR (see Figure 2). This system allows both interactive testing as well as long running experiments of RFID reading protocols. It comprises a software platform allowing fine control over all dynamic aspects influencing RFID readings: movements for target and antenna, RFID reader configuration, and smart antenna configuration (diversity and power control).





#### 5.3. Metamorphic Housing platform and Software - On-demand room

KEYWORDS: Smart Home - Metamorphic House - Sharing

- Participants: Ghislain Nouvel (Univ. Rennes 1), Guillermo Andrade Barroso and Michele Dominici
- Partner: Univ. Rennes 1
- Partner: Delta Dore Néotoa<sup>0</sup>
- Contact: Michele Dominici and Frédéric Weis

SCIENTIFIC DESCRIPTION- SOFTWARE

As part of the experimentation of the On-demand room, we have developed a software system that will be used to manage the room and provide functionalities to end users and building managers (access control, electrical and time consumption monitoring and report, room state display...). The software has been deployed in the building that hosts the experimentation. This software is co-developed by Michele Dominici (Univ. Rennes 1), Guillermo Andrade (SED Inria) and Ghislain Nouvel (MobBI platform <sup>0</sup>). Contributions are provided by members of the Diverse project-team. Intellectual protection is expected to be applied on such software.

#### SCIENTIFIC DESCRIPTION- PLATFORM

We realized a prototype of the on-demand room as an immersive interactive virtual-reality application, leveraging the Immersia platform (see https://raweb.inria.fr/rapportsactivite/RA2015/tacoma/uid29.html), with real domestic appliances connected to Immersa. In 2016-2017, the experimentation of the On-demand room is organized in the following steps: modification of the original building to create a common, On-demand room between two apartments; deployment of the computer and hardware and software that we are developing; rental of the apartments to two households, for an estimated duration of one year. The building that hosted the experimentation is showed in Figure 3. During the rental of the apartments, data has been collected and stored about the use of the room by households. Data included time of occupation, mode (private or shared),

<sup>&</sup>lt;sup>0</sup>http://www.neotoa.fr/

<sup>&</sup>lt;sup>0</sup>https://mobbi.univ-rennes1.fr/

consumptions, errors etc. The On-demand room thus constitutes an experimentation platform, where real people live and produce data that can be analyzed for statistical purposes. Produced data could also be used in combination with interviews of the occupants to improve the functionalities of the On-demand room, evaluate acceptance and appropriation.



Figure 3. On-demand room real experimentation

#### 5.4. ISO/IEC 15118-2 Open source Implementation

KEYWORDS: Smart Grid - Intelligent Transport System

- Partner: IMT Atlantique
- Participants: Guillaume Le Gall
- Contact: Jean-Marie Bonnin

SCIENTIFIC DESCRIPTION

The ISO/IEC 15118 standard, named "Road vehicles – Vehicle-to-Grid Communication Interface", defines how an electric vehicle and a charging station should communicate. It enables the Smart Charging of electric vehicles by allowing them to plan their charging sessions. As we want to be able to manage the charge of electric vehicles in our micro Smart Grid systems, we decided to implement the protocol defined by this standard. The goal is also to participate actively in the design of the new version of this protocol. During a charging session the charging station provides the vehicle with the status of the electric power grid. The vehicle is then able to plan its sharing session accordingly. It sends back its charge plan to the charging station, so that the Smart Grid is aware of it. The protocol also provides security and authentication features.

This software platform was implemented onto small PCs, and was used to control the charge in a small and portable demonstration platform, to demonstrate how it is possible to interconnect this high level decision and communication software with low level components, such as a Battery Management System (BMS), and a battery charger. In 2016, in the context of the Greenfeed project our software has been demonstrated to control the charge of the electric vehicle during the final demonstration of the project. The integration work has been done in collaboration with VeDeCom<sup>0</sup>.

# 6. New Results

#### 6.1. Smart City and ITS

**Participants:** Indra Ngurah, Djibrilla Amadou Kountche, Xavier Gilles, Christophe Couturier, Rodrigo Silva, Frédéric Weis, Jean-Marie Bonnin [contact].

<sup>&</sup>lt;sup>0</sup>http://www.vedecom.fr/

The domain of Smart Cities is still young but it is already a huge market which attract number of companies and researchers. It is also multi-fold as the words "smart city" gather multiple meanings. Among them one of the main responsibilities of a city, is to organisation the transportation of goods and people. In intelligent transportation systems (ITS), ICT technologies have been involved to improve planification and more generally efficiency of journeys within the city. We are interested in the next step where efficiency would be improved locally relying on local interactions between vehicles, infrastructure and people (smartphones).

For the future "autonomous" vehicle are now in the spotlight, since a lot of works has been done in recent years in automotive industry as well as in academic research centers. Such unmanned vehicle could strongly impact the organisation of the transportation in our cities. However, due to the lack of a definition of what is an "autonomous" vehicle it remains still difficult to see how these vehicles will interact with their environment (eg. road, smart city, houses, grid, etc"). From augmented perception to fully cooperative automated vehicle, the autonomic cover various realities in terms of interaction the vehicle relies on. The extended perception relies on communication between the vehicle and surrounding roadside equipments. This help the driving system to build and maintain an accurate view of the environment. But at this first stage the vehicle only uses its own perception to make its decisions. At a second stage, it will take benefits of local interaction with other vehicles through car-to-car communications to elaborate a better view of its environment. Such "cooperative autonomy" does not try to reproduce the human behavior anymore, it strongly rely on communication between vehicles and/or with the infrastructure to make decision and to acquire information on the environment. Part of the decision could be centralized (almost everything for an automatic metro) or coordinated by a roadside component. The decision making could even be fully distributed but this put high constraints on the communications. Automated vehicles are just an exemple of smart city automated processes that will have to share information within the surrounding to make their decisions.

We participated in the definition of the distributed architecture that has been adopted by all partners of the SEAS project. The main principles of this architecture have been published and we developed several profs of concept that have been demonstrated in the project consortium. Our partner developed the components of the architecture that has been demonstrated in the final review of the project (in January). The principles of the architecture and data representation has been used to design an open reusable Data Manager in the context of the EkoHub projet. This modular software will be extended to fit the needs of Indra Ngurah and Rodrigo Silva works.

#### 6.2. Convergence middleware for pervasive data

Participants: Yoann Maurel, Jules Desjardin, Paul Couderc [contact].

We are currently working on data driven middleware approaches dedicated to physical objects and smart spaces. We had previous contributions on the topic, where opportunistic collaborations between mobile devices were supported by Linda-like tuple space and IEEE 802.11 radios. However, these were adapted to relatively complex devices and the technological limitation at the time did not allow the full potential of the model. More recently, we investigated distributed storage spread over physical objects or fragments using RFID, enabling complex data to be directly reflected by passive objects (without energy). Yet other radio technologies, such as BLE, are emerging to support close range interactions with very low (or even zero) energy requirements.

Applications such as pervasive games (for ex. Pokemon Go), on the go data sharing, collaborative mobile app are often good candidates for opportunistic or dynamic interaction models. But they are not well supported by existing communication stacks, especially in context involving multiple technologies. Technological heterogeneity is not hidden, and high level properties associated with the interactions, such as proximity/range, or mobility-related parameters (speed, discovery latency) have to be addressed in an ad hoc manner. We think that a good way to solve these issues is to offer an abstract interaction model that could be mapped over the common proximity communication technologies, in a similar way as MOM (Message Oriented Middleware) such as MQTT abstract communications in many IoT and pervasive computing scenarios. However, they typically requires IP level communication, which far beyond the capabilities of ultra low energy proximity communication such as RFID and BLE. Moreover, they often rely on a coordinator node that is not adapted in highly dynamic context involving ephemeral communications and mobile nodes. We started the implementation of an associative memory mechanism over BLE, as it is a common ground that can be shared with passive or semi passive communications (RFID, NFC). Such mechanism, although relatively low level, is still a very useful building block for opportunistic applications: it enables opportunistic data storage/sharing and signaling/synchronization (in space in particular). This approach is fully in line with more general trend developed in the team to build "smart" systems leveraging local resources and data oriented mediation. We have started validation work with a few applications, in particular regarding energy aspects and scalability with respect to the communication load. This should lead to publishing on both infrastructure and application level aspects of the approach.

# 6.3. Modeling activities and forecasting energy consumption and production to promote the use of self-produced electricity from renewable sources

Participants: Alexandre Rio, Yoann Maurel [contact].

This work began in 2017 and is carried out as part of a broader collaboration between Tacoma, the Diverse Team and OKWind, a company specialized in the production of renewable sources of energy. OKWind proposes to deploy self-production units directly where the consumption is done. It has developed expertise in vertical-axis wind turbines, photovoltaic trackers, and heat pump. This project aims at building a system that optimizes the use of different sources of renewable energy, choosing the most suitable source for the current demand and anticipating future needs. The goal is to favor the consumption of locally produced electricity and to maximize the autonomy of the equipped sites so as to reduce the infrastructure needed to distribute electricity, to set energy cost, and to reduce the ecological impact of energy consumption.

Modeling and forecasting production and consumption of a site is hard and raises several issues: how to precisely assess the consumption and production of energy on a given site with changing conditions ? How to adequately size energy sources and energy storage (wind turbine, solar panel and batteries) ? And what methods to use to optimize consumption and, whenever possible, act on installations and activities to reduce energy costs ? We aim to propose tools to predict the consumption of a site based on estimation and previous observation, monitor the site in real time and forecast evolution. We propose to build a DSL describing consumption and production processes, and a system providing recommendations based on the derived model at runtime.

The problem of forecasting is known from both a production and consumption point of view. OKWind has developed tools to predict the production of their renewable sources - the same goes for batteries - and a lot of theoretical work has been done on consumption in the literature. In our view little has been done to precisely model activities, their energy consumption and the associated variability. Indeed most of the current approaches are concerned with either large-scale forecasting for the Smart Grid, are based on coarse grain data (total energy consumption of the site), or focuses on modeling specific appliance without describing how and when they are used.

This is paradoxical considering that companies have spent a lot of time modeling their activities from a logistic point of view. Intuitively, the predictable and seasonal nature of a company's activities would allow building activity schedulers that favor the consumption of certain energy sources (the cleanest or least expensive one for instance). The development of a DSL to describe the relationships between activities, their planning, and the production and environmental factors would make possible to simulate a given site at a given location, to make assumptions on sizing, and would be a basis to forecast energy consumption so as to provide recommendations for the organization of activities.

We already have developed part of this DSL that simulates activities and production. In particular, it is capable of simulating consumption and production over a given period based on available environmental data. This tool is in the experimentation phase. In particular, we are collecting information on several sites to measure the consumption of various activities.

#### 6.4. Sharing knowledge and access-control

Participants: Adrien Capaine, Yasmina Andaloussi, Frédéric Weis, Yoann Maurel [contact].

Smart spaces (Smart-city, home, building, etc.) are complex environments made up of resources (cars, smartphones, electronic equipment, applications, servers, flows, etc.) that cooperate to provide a wide range of services to a wide range of users. They are by nature extremely fluctuating, heterogeneous, and unpredictable. In addition, applications are often mobile and have to migrate or are offered by mobile platforms such as smartphones or vehicles.

To be relevant, applications must be able to adapt to users by understanding their environment and anticipating its evolutions. They are therefore based, explicitly or implicitly, on a representation of their surrounding environment based on available data provided by sensors, humans, objects and applications when available. The accuracy of the adaptations made by the applications depends on the precision of this representation. Building and maintaining such knowledge is resource-intensive in terms of network exchanges, computing time and incidentally energy consumption. It is, therefore, crucial to find ways to improve this process. In practice, many applications build their own models without sharing them or delegating calculations to remote services, which is not optimal because many processes are redundant. A huge improvement would be to find mechanisms that allows sharing the information so as to reduce as much as possible the treatments necessary to obtain it.

However, it seems extremely complex to provide a global, complete and unified view of the environment that reflects the applications' concerns. If it were possible, such a single representation would by nature be incomplete or subjective. Our solution should be applicable to nowadays devices and applications with little adjustments to the underlying architectures. It should then be flexible enough to deal with the lack of standards in the domain without imposing architectural choices. Such lack of standard is very common in IT and mainly due to well-known factors: (1) for technical reasons, developers tend to think that their "standard" is better suited for their current use-case, or/and (2) for commercial reasons companies want to keep a closed siloed system to capture their users, or/and (3) because the domain is still new and evolving and no standard as emerged yet, or/and finally (4) because the problem is too complex to be standardized and most proposed standards tend to be bloated and hard to use. The IoT domain suffers from all of these impediments and solution targeting mid-term application have to take these factors into accounts. Many IoT applications are still organized in silos of information. This leads to the deployment of sensors with similar functions and redundant pieces of software providing exactly the same service. Many frameworks or ontologies have been developed in the field to provide a solution to this problem but their implementation depends on the goodwill of the companies who do not always see their interest in losing part of the control of their application and data. To be largely accepted, solutions should let companies decide what information to share and when with little impact on their current infrastructure.

We want to be able to develop collaborative mechanisms that allow applications to share some of their information about the immediate surrounding environment with their counterparts. The idea is to allow the construction of shared representations between groups of applications that manipulate the same concepts so that each group can construct a subjective and complete representation of the environment that corresponds to its concerns. In this context, we want to offer applications mechanisms allowing them to leave information about their environment by associating them directly with the flows, data, services and objects handled. This information will be stored by the environment so that it will be possible for the application to retrieve it and for its peers to access it. From a logical point of view, applications will have the illusion of annotating objects directly; we make no assumptions about where this information will be stored as close as possible to the environments they qualify for reasons of performance, confidentiality and autonomy. To experience that idea, we have developed:

• Matriona, a globally distributed framework developed on top of OSGi. This project has been described in more details in the previous activity report. It is meant to be a global framework for exposing devices as REST-like resources. Resources functionalities can be extended through the mean of decorators. The system also provides access mechanisms. The main interest of Matriona

with regards to the information enrichment is its ability to support the dynamic extension of resource meta-information by application and to provide means to share this meta-information with others. It implements the concept of groups of interest with access control on meta-information. The concept described in Matriona are in the process to be published.

• Little Thumb Base (LithBase) is an independent knowledge base that provides the same enrichment capabilities than Matriona but imposes fewer constraints on the architecture of applications. It is a shared database implemented on simple low power nodes (esp32) that are cheap to deploy, flash and use. The idea behind LithBase is to decouple the storage from the framework and to provide a standard mechanism to share information. Ultimately we want to use its capabilities to implement a registry in the manner of Consul with meta-information enrichment and sharing mechanisms. By focussing only on the discovery mechanism and information sharing, LithBase imposes fewer constraints on applications and comply more with the goal of being ready to use in existing applications. This is still a work in progress. This solution also raises the issue of trust and control over access to this information. It is indeed necessary for applications to be able to determine the source of the additional information and to determine who will have access to the information they add. We have also been experimenting with access control mechanism that is implemented by LithBase. We are currently using elliptic cryptography to allow private information sharing between groups. Ultimately the goal of this project is to produce a coordinating object that implements generic mechanisms favouring opportunistic behaviours of surrounding applications.

# 7. Bilateral Contracts and Grants with Industry

#### 7.1. Bilateral Contracts with Industry

#### **Project: SIMHet**

Partner: YoGoKo

Starting: Nov 2015; Ending : October 2018

#### Contact: JM Bonnin

Abstract: The SIMHet project is performed in partnership with YoGoKo, a start-up that develops innovative communication solutions for cooperative intelligent transport systems. The SIMHet project aims to develop a decision making mechanism that would be integrated in the ISO/ETSI ITS communication architecture. It will allow mobile devices or mobile routers to choose the best network interface for each embedded application/flow. For example, in a vehicular environment this mechanism could manage global (Internet) and local connections for each on board device/application, in order to ensure that applications and services are always best connected. Aware that "best" concept is context-dependent, such a decision making mechanism should take into account requirements from different actors (e.g., applications, user, network administrators) and contextual information. One of the difficulties is to take advantage of the knowledge the system could have about near future connectivity. In the vehicular context such information about the movement and the availability of network resources is available. If taking into account the future makes the decision making more complex, this could allow a better usage of network resources when they are available. Once current solutions in the market are based on very simple decisions (use WiFi if available and 3G elsewhere), this smart mechanism will give competitive advantage for YoGoKo over its competitors.

# 8. Partnerships and Cooperations

#### 8.1. Regional Initiatives

**Project: EkoHub** 

Partner: Ekolis, Delaye transport, Telecom Bretagne

Starting: Nov 2014; Ending : Nov 2017

Contact: JM Bonnin

Abstract: The EkoHub project has been architectured around our multi-technologies gateway and leverages on the one developed in the ITSSv6 European project. In addition to the multiple interfaces of our platforms, sensor devices have been incorporated into the project and we studied different scenarios elaborated with our professional partners (Layaye Logistics). Intelligent data management schemes are being studied to adapt to the communication environment and the needs of the application consuming the data. The data model has been derived from the outcomes of the SEAS project.

The final EkoHub demonstration held in november 2017 with project partners.

#### 8.2. European Initiatives

#### 8.2.1. Collaborations in European Programs, Except FP7 & H2020

#### Project: SCOOP@F part 2

Partner: MEDE, Renault, PSA

Starting: Jan 2016; Ending : Dec 2018

Coordinator: JM Bonnin

Abstract: SCOOP@F is a Cooperative ITS pilot deployment project that intends to connect approximately 3000 vehicles with 2000 kilometers of roads. It consists of 5 specific sites with different types of roads: Ile-de-France, "East Corridor" between Paris and Strasbourg, Brittany, Bordeaux and Isère. SCOOP@F is composed of SCOOP@F Part 1 from 2014 to 2015 (ongoing) and SCOOP@F Part 2 from 2016 to 2018. Its main objective is to improve the safety of road transport and of road operating staff during road works or maintenance. The project includes the validations of Cooperative ITS services in open roads, cross border tests with other EU Member States (Spain, Portugal and Austria) and development of a hybrid communication solution (3G-4G/ITS G5). We are involved in the project to study the security and privacy properties of the hybrid architecture that allow to use non dedicated communication networks (WiFi, 5G) as well as the vehicular dedicated communication technologies (G5).

#### Project acronym: SEAS (ITEA3)

Partners: Telecom Paris Tech, Telecom Saint Etienne, Mines Saint Etienne, Engie, Kerlink, BeNomad, ICAM, CNR, VTT

Starting: Feb 2014; ending: Jan 2017

Contact: JM Bonnin

Abstract: The SEAS project addresses the problem of inefficient and unsustainable energy consumption, which is due to a lack of sufficient means to control, monitor, estimate and adapt the energy use of systems versus the dynamic use situations and circumstances influencing the energy use. The objective of the SEAS project is to enable energy, ICT and automation systems to collaborate at consumption sites, and to introduce dynamic and refined ICT-based solutions to control, monitor and estimate energy consumption. Proposed solution should enable energy market participants to incorporate micro-grid environments and active customers. We are involved in the project to design a distributed system architecture and to implement two proofs of concept: the first one is related to the electric vehicle charging and the other one to the prevision of solar energy production.

#### **Project: SCHIEF**

Partner: TUM (Technical University of Munchen), IMT Atlantique, Eurecom Starting: Sept 2016; Ending : Dec 2018

#### Coordinator: JM Bonnin

Abstract: In SCHEIF, we create a pilot for an enabler platform for the industrial Internet of Things. We envision a three-layered architecture with Sensors and actuators on the lowest layer. This layer includes industrial robots. On top of this hardware layer we envision site-local processing of data. Such a processing is beneficial since it allows keeping latency boundaries on the one hand and being in full control of all data on the other hand. The latency is relevant for enabling diverse time-critical operations as they often happen in industrial production environments. The local processing is relevant for protecting data. A privacy-conform processing is required to protect company secrets and to protect the privacy of workers. The third layer comprises data processing in the cloud. We envision mostly local data processing. However, offloading computing tasks to public or private clouds will be relevant for compute-intense tasks and those tasks that require coordination between production sites. The main scenario of SCHEIF is an industrial production site where mobile robots and human workers coexist. The focus is providing the data required to manage and optimize the production process always at the most suitable quality. The suitability of data relies on the requirements of the data producers and consumers. A planned demo scenario is a provoked system crash that leads to reprioritization of data streams to mitigate from the failure.

# 9. Dissemination

#### 9.1. Promoting Scientific Activities

#### 9.1.1. Scientific Events Organisation

9.1.1.1. Member of the Organizing Committees

• Co-organizer of a local IoT conference, Technoconference 22, IoT et si on connectait le monde: enjeux d'aujourd'hui et de demain, October 5th 2017 - Rennes, P. Couderc.

#### 9.1.2. Scientific Events Selection

- 9.1.2.1. Reviewer
  - Mobile Networks and Applications, JM. Bonnin
  - IEEE transactions on Mobile Computing, JM. Bonnin
  - Simulation Modeling Practice and Theory, JM. Bonnin
  - Computer Standards & Interfaces, an international journal on Engineering Science and Technology, JM. Bonnin

#### 9.1.2.2. Member of the Conference Program Committees

• PC member for VTC, NGNS, CARI, ICON, ICIN, JM Bonnin

#### 9.1.3. Invited Talks

- Invited talk at DATE 2017: "Which interaction models between objects for smarter spaces?", JM Bonnin
- Invited talk at ENOVA (16/03/2017) Strasbourg: "BMS et charge intelligente de véhicule électrique: une introduction au protocole ISO/IEC 15118", JM Bonnin
- Invited tutoriel at Digicosme (Saclay) summer school: "Foster digital innovation in Smart Cities", JM Bonnin
- Invited talk at UTS (22/08/2017): "Fostering digital innovation in Smart Cities and Cooperative autonomous vehicle", JM Bonnin
- Invited talk at Technoconference, IoT et si on connectait le monde: enjeux d'aujourd'hui et de demain: "Développement d'applications pervasives dans un environnement IoT", F. Weis

#### 9.1.4. Leadership within the Scientific Community

- Jean-Marie Bonnin is member of the scientific council of the GIS ITS
- Jean-Marie Bonnin is member of the scientific council the Id4Car cluster
- Jean-Marie Bonnin is an elected member of the Scientific Council of IMT
- Jean-Marie Bonnin is a scientific advisor of the YoGoKo startup

#### 9.1.5. Scientific Expertise

- Evaluation committee for the Belgium government, JM. Bonnin
- Expert for ANR, F. Weis and JM Bonnin
- Evaluation committee for the CADO (PME Business) challenge (Id4Car, I&R), JM Bonnin
- Expert for CSV board of "Pôle Images et Réseaux", P. Couderc

#### 9.1.6. Research Administration

• Head of the Networks, Telecommunication and service department at IRISA, JM. Bonnin

#### 9.2. Teaching - Supervision - Juries

#### 9.2.1. Teaching

L2/L3: network computing (lectures, tutorials, labs), 250 hours, F. Weis, Univ. Rennes 1

Master: Wireless LANs, F. Weis, 20 hours, M2, IMT Atlantique

Master: Supervision of a Master 1 project related to the smart building (in collaboration with Myriads), 22 hours, P. Couderc, Univ. Rennes 1

Master 1: Network programming (lectures, tutorial, labs), 78 hours, Y. Maurel

L3/M2: network communications protocol for building automation (lectures, labs), 80 hours, Y. Maurel

Master 2: Software engineering (lectures, tutorial, labs), 82 hours, Y. Maurel

#### 9.2.2. Supervision

PhD in progress: Adrien Capaine, Vers une plate-forme de LED connectées comme vecteur de services contextuels dans le cadre des bâtiments intelligents, 01/05/15, Frédéric Weis and Yoann Maurel

PhD in progress: Zaineb Lioune, Une Architecture pour des Services e-Santé évolutifs dans le cadre des Maisons Intelligentes, 01/09/14, Frédéric Weis, Tayeb Lemlouna and Philippe Roose

PhD in progress: Indra Ngurah, Car-based Data Collection for Low Energy Devices (Car-based DC4LED), 01/05/16, Jean-Marie Bonnin

PhD in progress: Christophe Couturier, Frugal networking for cooperative autonomy, 01/11/16, Jean-Marie Bonnin

PhD in progress: Alexandre Rio, Modélisation des activités de site consommateur d'énergie, 01/10/16, Olivier Barais and Yoann Maurel

PhD in progress: Rodrigo Silva, Mécanisme de décision multi-critères pour le placement de flux en environnement hétérogène et changeant, 1/11/15, Jean-Marie Bonnin

#### 9.2.3. Juries

Adrien Carteron, "Une approche événementielle pour le développement de services multi-métiers dédiés à l'assistance domiciliaire", Univ. Bordeaux, F. Weis, PhD referee

Amina BOUDENDIR, "Virtual Network Functions modeling for Dynamic Network-as-a-Service: Architecture principles and deployment", JM Bonnin, PhD referee

Ikbel Daly, JM Bonnin, PhD referee

#### 9.3. Popularization

Under the leitmotif "testing and learning about digital mobility solutions in Rennes", Rennes Metropole will open the doors of its city from March 14 - 18, 2018 for a new type of inclusive and collaborative event experience, dedicated to transforming mobility services: inOut! Composed of two distinct but complementary parts, inOut will combine debates among professionals from established and incumbent companies, start-ups, academics, students and live experiences on the city territory including the citizens of Rennes and its surroundings. More than just a recurrent annual event, inOut is set to continue throughout the year as a unique incubator on a human scale, where new mobilities are invented [IN]door and are tested [OUT]doors with users in real time. It's a one-of-a-kind co-creation platform providing industrial partners, start-ups, incumbent and new mobility players the opportunity to invent new mobility services, to experiment and test new use cases and business models, in close proximity to and with the citizens. Through a series of controversial debates, the event invites professionals, academics and local authorities from all over the world to separate hype from reality and discuss enablers and potential blocking points of these new mobilities.

We have been strongly involved in the creation of InOut and Jean-Marie Bonnin is responsible of the "Expert Committee" of the event.

This autumn the work done to design the InOut concept provided the basis to build the answer of Rennes Metropole to the AMI "Territoire Innovant de Grande Ambition".

# **10. Bibliography**

#### **Publications of the year**

#### Articles in International Peer-Reviewed Journal

- [1] Z. LIOUANE, T. LEMLOUMA, P. ROOSE, F. WEIS, M. HASSANI. An improved extreme learning machine model for the prediction of human scenarios in smart homes, in "Applied Intelligence", September 2017 [DOI: 10.1007/s10489-017-1062-5], https://hal.inria.fr/hal-01678671.
- [2] S. MULLER, C. HARPES, Y. LE TRAON, S. GOMBAULT, J.-M. BONNIN. Efficiently computing the likelihoods of cyclically interdependent risk scenarios, in "Computers and Security", January 2017, vol. 64, p. 59 - 68 [DOI: 10.1016/J.COSE.2016.09.008], https://hal.archives-ouvertes.fr/hal-01427488.

#### **International Conferences with Proceedings**

- [3] S. EL JAOUHARI, A. BOUABDALLAH, J.-M. BONNIN, T. LEMLOUMA.Securing the Communications in a WoT/WebRTC-based Smart Healthcare Architecture, in "MT4H 2017 : 2nd International Workshop on Mobile Technology for Healthcare", Exeter, United Kingdom, IEEE, June 2017, vol. I-SPAN-FCST-ISCC 2017, p. 403 - 408 [DOI: 10.1109/ISPAN-FCST-ISCC.2017.70], https://hal.archives-ouvertes.fr/hal-01666280.
- [4] S. EL JAOUHARI, A. BOUABDALLAH, J.-M. BONNIN, T. LEMLOUMA.*Toward a Smart Health-care Architecture Using WebRTC and WoT*, in "WorldCIST 2017", Porto Santo Island, Madeira, Portugal, Á. ROCHA, A. M. CORREIA, H. ADELI, L. P. REIS, S. COSTANZO (editors), Recent Advances in Information Systems and Technologies, Springer, April 2017, vol. Advances in Intelligent Systems and Computing, n<sup>o</sup> 571, p. 531-540 [DOI: 10.1007/978-3-319-56541-5\_54], https://hal.archives-ouvertes.fr/hal-01503434.

#### Scientific Books (or Scientific Book chapters)

[5] S. EL JAOUHARI, A. BOUABDALLAH, J.-M. BONNIN. Security issues of the web of things, in "Managing the Web of Things : Linking the Real World to the Web", Elsevier, 2017, p. 389 - 424 [DOI: 10.1016/B978-0-12-809764-9.00018-4], https://hal.archives-ouvertes.fr/hal-01511323.

#### **References in notes**

[6] R. BALDONI, M. BANÂTRE, M. DOMINICI, D. DUMORTIER, M. EL MANKIBI, M. FRÉJUS, B. KOLDE-HOFE, M. MECELLA, K. ROTHERMEL, P. SUIGNARD. Comfort-Oriented Metamorphic House (COMETAE), October 2012, n<sup>o</sup> PI 1998, Publications Internes de l'IRISA ISSN : 2102-6327, https://hal.inria.fr/hal-00864884.

# **Team TAMIS**

# Threat Analysis and Mitigation for Information Security

Inria teams are typically groups of researchers working on the definition of a common project, and objectives, with the goal to arrive at the creation of a project-team. Such project-teams may include other partners (universities or research institutions).

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Security and Confidentiality

## **Table of contents**

1.	Personnel						
2.	Overall Objectives						
	2.1. Context	934					
	2.2. Approach and motivation	935					
3.	Research Program	935					
	3.1. Axis 1: Vulnerability analysis	935					
	3.2. Axis 2: Malware analysis	936					
	3.3. Axis 3: Building a secure network stack	936					
4.	Application Domains	<b>937</b>					
	4.1. System analysis	937					
	4.2. Cybersecurity	937					
	4.3. Safe Internet	937					
5.	Highlights of the Year	<b>937</b>					
6.	New Software and Platforms	<b>938</b>					
	6.1. GNUnet	938					
	6.2. MHD	938					
	6.3. PLASMA Lab	938					
	6.4. Taler	939					
	6.5. HyLeak	939					
	6.6. SimFI	940					
	6.7. DaD	940					
	6.8. MASSE	940					
	6.9. Behavioral Malware Analysis	940					
	6.10. VITRAIL - Visualisation Tool	941					
	6.11. VITRAIL 6 JBInsTrace	941					
	6.12. Platforms	941					
	6.12.1. Malware'o'Matic	941					
	6.12.2. Faustine	942					
7.	New Results	942					
	7.1. Results for Axis 1: Vulnerability analysis	942					
	7.1.1. Statistical Model Checking of LLVM Code	942					
	7.1.2. Verification of IKEv2 protocol	942					
	7.1.3. High-Level Frameworks for Scheduling Systems	943					
	7.1.4. Side-channel Analysis of Cryptographic Substitution Boxes	943					
	7.1.5. New Advances on Side-channel Distinguishers	945					
	7.1.6. Side-channel analysis on post-quantum cryptography	948					
	7.1.7. Binary Code Analysis: Formal Methods for Fault Injection Vulnerability Detection	948					
	7.1.8. Security at the hardware and software boundaries	949					
	7.1.8.1. Side-channel attacks (SCA)	949					
	7.1.8.2. SCA based fuzzer	949					
	7.1.9. System Vulnerability Analysis	950					
	7.1.9.1. Reverse engineering	950					
	7.1.9.2. Safe system update mechanism	950					
	7.1.9.3. Prediction of system divergence	951					
	7.2. Results for Axis 2: Malware analysis	951					
	7.2.1. Malware Detection	951					
	7.2.2. Malware Deobfuscation	952					
	7.2.3. Malware Classification	953					
	7.2.4. Botnet Trojan Detection	953					

	7.2	.5.	Modu	alar Automated Syntactic Signature Extraction (MASSE)	954		
	7.2	.6.	Malw	vare IDS	954		
	7.2.7.			S	955		
	7.3.	Resu	ilts fo	r Axis 3: Building a secure network stack	956		
	7.3.1.		Priva	cy-Preserving Abuse Detection in Future Decentralised Online Social Networks	956		
	7.3.2.		Fog c	of Trust	956		
	7.3.3.		Cell t	ower privacy	956		
	7.3.4.		Taler	protocol improvements	957		
	7.3.		Mix 1	Networking	957		
	7.4.	Othe	er rese	arch results	957		
7.4.1. Pri			Priva	cy and Security: Information-Theoretical Quantification of Security Properties	957		
	7.4	.2.	2. Security for therapeutical environments				
	7.4.3. Mobile air pollution sensor platform for smart-cities				961		
8.	Bilate	ral C	ontra	cts and Grants with Industry	. 961		
	8.1. Bilateral Contracts with Industry				961		
	8.2.	Bila	teral (	Grants with Industry	961		
9.	Partn	ershi	ps an	d Cooperations	. 961		
	9.1.	Regi	ional l	Initiatives	961		
	9.2.	Nati	onal I	nitiatives	962		
	9.2	.1.	ANR		962		
	9.2	.2.	DGA		962		
	9.2	.3.	Autre	S	962		
	9.3.	Euro	opean	Initiatives	962		
		9.3.1	l.1.	ACANTO	962		
		9.3.1	.2.	DIVIDEND	963		
		9.3.1	1.3.	EMC2	964		
		9.3.1	.4.	ENABLE-S3	965		
		9.3.1	.5.	SUCCESS	967		
		9.3.1	.6.	TeamPlay	967		
10.	Disse	emina	ation		. 968		
	10.1. Promoting Scientific Activities				968		
	10.	1.1.	Scien	tific Events Organisation	968		
		10.1	.1.1.	General Chair, Scientific Chair	968		
		10.1	.1.2.	Member of the Organizing Committees	968		
	10.	1.2.	Scien	tific Events Selection	968		
		10.1	.2.1.	Member of Conference Steering Committees	968		
		10.1	.2.2.	Chair of Conference Program Committees	968		
		10.1	.2.3.	Member of the Conference Program Committees	969		
		10.1	.2.4.	Reviewer	969		
	10.	1.3.	Journ	al	969		
		10.1	.3.1.	Member of the Editorial Boards	969		
		10.1	.3.2.	Reviewer - Reviewing Activities	969		
	10.	1.4.	Invite	ed Talks	969		
	10.	1.5.	Scien	tific Expertise	969		
	10.	1.6.	Resea	arch Administration	970		
	10.2. Teaching - Supervision - Juries						
	10.	2.1.	Teach	ling	970		
	10.2.2. Supervision				970		
	10.2.3. Juries						
	10.3.	Рори	ulariza	ation	971		
11.	Bibli	iogra	phy		. <b>971</b>		

#### **Team TAMIS**

*Creation of the Team: 2016 January 01, updated into Project-Team: 2018 January 01* **Keywords:** 

#### **Computer Science and Digital Science:**

- A4. Security and privacy
- A4.1. Threat analysis
- A4.2. Correcting codes
- A4.4. Security of equipment and software
- A4.5. Formal methods for security
- A4.8. Privacy-enhancing technologies
- A4.9. Security supervision

#### **Other Research Topics and Application Domains:**

- B6. IT and telecom
- B6.4. Internet of things
- B6.5. Information systems
- B6.6. Embedded systems
- B8.1. Smart building/home
- B8.2. Connected city
- B8.4. Security and personal assistance
- B9.8. Privacy
- B9.9. Risk management
- B9.10. Ethics

# 1. Personnel

#### **Research Scientists**

Axel Legay [Team leader, Inria, Researcher, HDR] Christian Grothoff [Inria, Advanced Research Position, until Aug 2017, HDR] Annelie Heuser [CNRS, Researcher] Jean-Louis Lanet [Inria, Senior Researcher, HDR] Olivier Zendra [Inria, Researcher] Fabrizio Biondi [Centrale-Supelec, Researcher, "Chaire Malware"] Kim Larsen [Inria, International Chair, Advanced Research Position]

#### **External Collaborators**

Francois-Renaud Escriva [DGA] Sebastien Josse [DGA] Colas Le Guernic [DGA]

#### **Technical Staff**

Jeffrey Paul Burdges [Inria] Sébastien Campion [Inria] Thomas Given-Wilson [Inria] Bruno Lebon [Inria, from Oct 2017] Laurent Morin [Inria] Jean Quilbeuf [Inria] Sean Sedwards [Inria, until Mar 2017] Marcello Stanisci [Inria, until Jul 2017] Gabor Toth [Inria, until Apr 2017] Louis Marie Traonouez [Inria]

#### **PhD Students**

Delphine Beaulaton [UBS Vannes, from September 2017] Sebanjila Bukasa [Inria] Mounir Chadli [Algerian Ministry of Defense, until Dec 2017] Olivier Decourbe [Inria] Florian Dold [Inria] Mihai Enescu [Inria, until Sep 2017] Alexandre Gonzalvez [IMT Atlantique] Nisrine Jafri [Inria] Martin Moreau [SECURE-IC SAS, from May 2017] Ruta Moussaileb [IMT Atlantique, from Oct 2017] Tristan Ninet [Thales] Lamine Noureddine [Inria, from Nov 2017] Leopold Ouairy [Inria, from Oct 2017] Aurélien Palisse [Inria] Emmanuel Tacheau [CISCO, from Sep 2017] Alexander Zhdanov [Inria, from Oct 2017]

#### **Post-Doctoral Fellows**

Najah Ben Said [Inria] Ludovic Claudepierre [Inria, from Oct 2017] Ronan Lashermes [Inria, until Oct 2017] Hélène Le Bouder [Inria, until Mar 2017] Tania Richmond [Inria, from Sep 2017]

Administrative Assistant Cecile Bouton [Inria]

# 2. Overall Objectives

#### 2.1. Context

Security devices are subject to drastic security requirements and certification processes. They must be protected against potentially complex exploits that result from the combination of software and hardware attacks. As a result, a major effort is needed to develop new research techniques and approaches to characterize security issues, as well as to discover multi-layered security vulnerabilities in complex systems.

In recent years, we have witnessed two main lines of research to achieve this objective.

The first approach, often called *offensive security*, relies on engineering techniques and consists in attacking the system with our knowledge on its design and our past expertise. This is a creative approach that supports (1) checking whether a system is subject to existing vulnerabilities, i.e. classes of vulnerabilities that we already discovered on other systems, and (2) discovering new types of vulnerabilities that were not foreseen and that may depend on new technologies and/or programming paradigms. Unfortunately, this approach is limited to systems whose complexity remains manageable at the human level. This means that exploits which combine several vulnerabilities may be hard to identify. The second and more formal approach builds on formal models (also known as *formal methods*) to automatically detect vulnerabilities, or prove their absence. This is applicable to systems whose complexity is beyond human reasoning, but can only detect existing classes of vulnerabilities, i.e., those that have been previously characterized by offensive security.

#### 2.2. Approach and motivation

The claim made by TAMIS is that assessing security requires combining both engineering and formal techniques.

As an example, security exploits may require combining classes of well-known vulnerabilities. The detection of such vulnerabilities can be made via formal approaches, but their successful combination requires human creativity. TAMIS's central goal is thus to demonstrably narrow the gap between the vulnerabilities found using formal verification and the issues found using systems engineering. As a second example, we point out that there are classes of attacks that exploit both the software and hardware parts of a system. Although vulnerabilities can be detected via formal methods in the software part, the impact of attacking the hardware still needs to be modeled. This is often done by observing the effect of parameter changes on the system, and capturing a model of them. To address this situation, the TAMIS team bundled resources from scalable formal verification and secure software engineering for *vulnerability analysis*, which we extend to provide methods and tools to (a) *analyze (binary) code including obfuscated malware*, and (b) *build secure systems*.

Very concrete examples better illustrate the differences and complementarity of engineering and formal techniques. First, it is well-known that formal methods can be used to detect buffer overflows. However, the definition of buffer overflows itself was made first in 1972 when the Computer Security Technology Planning study laid out the technique and claimed that over sizing could be exploited to corrupt a system. This exploit was then popularized in 1988 as one of the exploits used by the Morris worm, and only at that point systematic techniques were developed to detect it. Another example is the work we conducted in attacking smart cards. The very firsts experiments were done at the engineering level, and consisted of retrieving the key of the card in a brute force manner. Based on this knowledge, we generated user test-cases that characterize what should not happen. Later, those were used in a fully automatized model-based testing approach [66].

# 3. Research Program

#### 3.1. Axis 1: Vulnerability analysis

This axis proposes different techniques to discover vulnerabilities in systems. The outcomes of this axis are (a) new techniques to discover system vulnerabilities as well as to analyze them, and (b) to understand the importance of the hardware support.

Most existing approaches used at the engineering level rely on testing and fuzzing. Such techniques consist in simulating the system for various input values, and then checking that the result conforms to a given standard. The problem being the large set of inputs to be potentially tested. Existing solutions propose to extract significant sets by mutating a finite set of inputs. Other solutions, especially concolic testing developed at Microsoft, propose to exploit symbolic executions to extract constraints on new values. We build on those existing work, and extend them with recent techniques based on dissimilarity distances and learning. We also account for the execution environment, and study techniques based on the combination of timing attacks with fuzzing techniques to discover and classify classes of behavior of the system under test.

Techniques such as model checking and static analysis have been used for verifying several types of requirements such as safety and reliability. Recently, several works have attempted to adapt model checking to the detection of security issues. It has clearly been identified that this required to work at the level of binary code. Applying formal techniques to such code requires the development of disassembly techniques to obtain a semantically well-defined model. One of the biggest issues faced with formal analysis is the state space explosion problem. This problem is amplified in our context as representations of data (such as stack content) definitively blow up the state space. We propose to use statistical model checking (SMC) of rare events to efficiently identify problematic behaviors.

We also seek to understand vulnerabilities at the architecture and hardware levels. Particularly, we evaluate vulnerabilities of the interfaces and how an adversary could use them to get access to core assets in the system. One particular mechanism to be investigated is the DMA and the so-called Trustzone. An ad-hoc technique to defend against adversarial DMA-access to memory is to keep key material exclusively in registers. This implies co-analyzing machine code and an accurate hardware model.

#### **3.2.** Axis 2: Malware analysis

Axis 1 is concerned with vulnerabilities. Such vulnerabilities can be exploited by an attacker in order to introduce malicious behaviors in a system. Another method to identify vulnerabilities is to analyze malware that exploits them. However, modern malware has a wide variety of analysis avoidance techniques. In particular, attackers obfuscate the code leading to a security exploit. For doing so, recent black hat research suggests hiding constants in program choices via polynomials. Such techniques hinder forensic analysis by making detailed analysis labor intensive and time consuming. The objective of research axis 2 is to obtain a full tool chain for malware analysis starting from (a) the observability of the malware via deobfuscation, and (b) the analysis of the resulting binary file. A complementary objective is to understand how hardware attacks can be exploited by malwares.

We first investigate obfuscation techniques. Several solutions exist to mitigate the packer problem. As an example, we try to reverse the packer and remove the environment evaluation in such a way that it performs the same actions and outputs the resulting binary for further analysis. There is a wide range of techniques to obfuscate malware, which includes flattening and virtualization. We will produce a taxonomy of both techniques and tools. We will first give a particular focus to control flow obfuscation via mixed Boolean algebra, which is highly deployed for malware obfuscation. We recently showed that a subset of them can be broken via SAT-solving and synthesis. Then, we will expand our research to other obfuscation techniques.

Once the malware code has been unpacked/deobfuscated, the resulting binary still needs to be fully understood. Advanced malware often contains multiple stages, multiple exploits and may unpack additional features based on its environment. Ensuring that one understands all interesting execution paths of a malware sample is related to enumerating all of the possible execution paths when checking a system for vulnerabilities. The main difference is that in one case we are interested in finding vulnerabilities and in the other in finding exploitative behavior that may mutate. Still, some of the techniques of Axis 1 can be helpful in analyzing malware. The main challenge for axis 2 is thus to adapt the tools and techniques to deal with binary programs as inputs, as well as the logic used to specify malware behavior, including behavior with potentially rare occurrences. Another challenge is to take mutation into account, which we plan to do by exploiting mining algorithms.

Most recent attacks against hardware are based on fault injection which dynamically modifies the semantics of the code. We demonstrated the possibility to obfuscate code using constraint solver in such a way that the code becomes intentionally hostile while hit by a laser beam. This new form of obfuscation opens a new challenge for secure devices where malicious programs can be designed and uploaded that defeat comprehensive static analysis tools or code reviews, due to their multi-semantic nature. We have shown on several products that such an attack cannot be mitigated with the current defenses embedded in Java cards. In this research, we first aim at extending the work on fault injection, then at developing new techniques to analyze such hostile code. This is done by proposing formal models of fault injection, and then reusing results from our work on obfuscation/deobfuscation.

#### **3.3.** Axis **3:** Building a secure network stack

To evaluate the techniques developed in Axes 1 and 2, we analyze concrete systems developed not only with industry partners, but also within the team. By using our own systems, we can co-evolve best-practices, while externally developed systems provide realistic challenges especially with respect to analyzing obfuscated malware in the hardware or complex vulnerabilities. In this context, Christian Grothoff (ARP Inria) is currently developing a new Internet, which is supposed to be more secure. This introduces interesting challenges both in terms of vulnerability and malware analysis, and hence should be a great opportunity to mix the competences of all the members of the team.
More precisely, this system intends to challenge the idea that network security is an administrative task, where network administrators shield users with passwords, firewalls, intrusion detection systems and policies. Instead, we want to eliminate administrators that have power over user's data, and as such administrators themselves are liabilities, and because a network design that permits administrative intrusion inherently adds vulnerabilities. Instead, the system should ensure secure communication mechanisms without trusted third parties.

Key challenges we work on include (a) improving scalable secure ad-hoc decentralized routing, including keyvalue lookup, unicast and multicast communication, (b) protecting meta-data in the overlay using advanced decentralized onion routing, (c) a unified public-key infrastructure and identity management solution that is suitable to replace the Web-of-Trust, X.509, DNSSEC and other legacy methods for naming and identifying services, (d) secure synchronous and asynchronous messaging at scale, providing decentralized alternatives to common online social applications and addressing challenges in protocol evolution and compatibility. Finally, we are currently working on GNU Taler, a new secure privacy-preserving payment system where users never have to authenticate. This system in particular can be used as a concrete test case for the methods developed in the team.

To support this research work, we develop a framework named GNUnet. It provides a clear separation into layers, which facilitates testing and verifying the various components. However, we see that often existing formal verification techniques still do not scale to typical subsystems encountered in practice. Our objective is thus to exploit efficient and scalable formal techniques techniques proposed in Axis 1 together with engineering skills in order to guide the validation (message synchronization, data protection, ...) and reach the best compromise. An additional complication is that we need a validation process that not merely covers the software itself, but also all of its dependencies (such as database, cryptographic libraries and networking libraries). For the Taler-specific hardware, we are envisioning an NFC-powered device, which creates new challenges in terms of securing cryptographic computations in a setting where the adversary has control over the power supply. In such a case, the attacker can drive the environment and modify the behavior of the system as we have shown in Axis 2. Providing the control of the environment is a new vector for attackers.

Christian Grothoff, who leads this axis, got a position in Bern in 2017. This axis is expected to follow him in the future, although Tamis still holds expertise and members to finish ongoing work with the team. Cooperations with Bern are expected in the future.

# 4. Application Domains

# 4.1. System analysis

The work performed in Axes 1 and 2 and the methods developed there are applicable to the domain of system analysis, both wrt. program analysis and hardware analysis.

# 4.2. Cybersecurity

The work done in the 3 axes above aims at improving cybersecurity, be it via vulnerability analyses, malware analyses and the development of safer networking mechanisms.

# 4.3. Safe Internet

The work done in Axis 3 above very directly contributes to the goal of a safer Internet.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

"Chaire Analyse de Menaces" (Threat Analysis)

#### Participants: Axel Legay, Fabrizio Biondi

Creation of the "Chaire Analyse de Menaces" (Threat Analysis), that has been assigned to Fabrizio Biondi.

## **Thales Air Operations partnership**

Participants: Axel Legay, Louis-Marie Traonouez

Creation of a partnership with Thales Air Operations for machine learning algorithms to detect anomalies in ground-to-air communications.

# 6. New Software and Platforms

# 6.1. GNUnet

SCIENTIFIC DESCRIPTION: The GNUnet project seeks to answer the question what a modern Internet architecture should look like for a society that care about security and privacy. We are considering all layers of the existing well-known Internet, but are also providing new and higher-level abstractions (such as voting protocols, Byzantine consensus, etc.) that are today solved in application-specific ways. Research questions include the desired functionality of the overall stack, protocol design for the various layers as well as implementation considerations, i.e. how to implement the design securely.

FUNCTIONAL DESCRIPTION: GNUnet is a framework for secure peer-to-peer networking that does not use any centralized or otherwise trusted services. Our high-level goal is to provide a strong free software foundation for a global network that provides security and in particular respects privacy.

GNUnet started with an idea for anonymous censorship-resistant file-sharing, but has grown to incorporate other applications as well as many generic building blocks for secure networking applications. In particular, GNUnet now includes the GNU Name System, a privacy-preserving, decentralized public key infrastructure.

- Participants: Alvaro Garcia Recuero, Florian Dold, Gabor Toth, Hans Grothoff, Jeffrey Paul Burdges and Sree Hrsha Totakura
- Partner: The GNU Project
- Contact: Hans Grothoff
- URL: https://gnunet.org/

# 6.2. MHD

#### GNU libmicrohttpd

KEYWORDS: Embedded - Web 2.0

SCIENTIFIC DESCRIPTION: We are providing a standards compliant and complete implementation of the HTTP server protocol that allows developers to easily write correct HTTP servers. Key challenges include code size minimization (for IoT devices), performance (zero copy, scalability to 100k concurrent connections), portability and security. MHD is already widely used in production by both academic and industrial users. Ongoing research challenges include formal verification.

FUNCTIONAL DESCRIPTION: GNU library that is supposed to make it easy to run an HTTP server as part of another application.

- Participants: Evgeny Grin, Hans Grothoff and Sree Hrsha Totakura
- Partner: The GNU Project
- Contact: Hans Grothoff
- URL: http://www.gnu.org/software/libmicrohttpd/

# 6.3. PLASMA Lab

KEYWORDS: Energy - Statistics - Security - Runtime Analysis - Model Checker - Statistical - Model Checking - Aeronautics - Distributed systems

SCIENTIFIC DESCRIPTION: Statistical model checking (SMC) is a fast emerging technology for industrial scale verification and optimisation problems. SMC only requires an executable semantics and is not constrained by decidability. Therefore we can easily apply it to different modelling languages and logics. We have implemented in PLASMA Lab several advanced SMC algorithms that combine formal methods with statistical tests, which include techniques for rare events estimation and non-deterministic models.

FUNCTIONAL DESCRIPTION: PLASMA Lab is a compact, efficient and flexible platform for statistical model checking of stochastic models. PLASMA Lab includes simulators for PRISM models (Reactives Modules Language-RML) and Biological models. It also provides plugins that interface external simulators in order to support Matlab/Simulink, SytemC and LLVM . PLASMA Lab can be extended with new plugins to support other external simulators, and PLASMA Lab API can be used to embed the tool in other softwares. PLASMA Lab provide fast SMC algorithms, including advanced techniques for rare events simulation and nondeterministic models. These algorithms are designed in a distributed architecture to run large number of simulations on several computers, either on a local area network or grid. PLASMA Lab is implemented in Java with efficient data structures and low memory consumption

NEWS OF THE YEAR: In 2017 we have extended PLASMA Lab with a new simulator plugin that allows to verify LLVM code.

- Participants: Axel Legay, Jean Quilbeuf, Benoît Boyer, Kevin Corre, Louis-Marie Traonouez, Matthieu Simonin and Sean Sedwards
- Contact: Axel Legay
- URL: https://project.inria.fr/plasma-lab/

# 6.4. Taler

#### GNU Taler

**KEYWORD:** Privacy

SCIENTIFIC DESCRIPTION: Taler is a Chaum-style digital payment system that enables anonymous payments while ensuring that entities that receive payments are auditable. In Taler, customers can never defraud anyone, merchants can only fail to deliver the merchandise to the customer, and payment service providers can be fully audited. All parties receive cryptographic evidence for all transactions, still, each party only receives the minimum information required to execute transactions. Enforcement of honest behavior is timely, and is at least as strict as with legacy credit card payment systems that do not provide for privacy.

The key technical contribution underpinning Taler is a new refresh protocol which allows fractional payments and refunds while maintaining untraceability of the customer and unlinkability of transactions. The refresh protocol combines an efficient cut-and-choose mechanism with a link step to ensure that refreshing is not abused for transactional payments.

We argue that Taler provides a secure digital currency for modern liberal societies as it is a flexible, libre and efficient protocol and adequately balances the state's need for monetary control with the citizen's needs for private economic activity.

FUNCTIONAL DESCRIPTION: Taler is a new electronic payment system. It includes an electronic wallet for customers, a payment backend for merchants and the main payment service provider logic called the exchange. Taler offers Chaum-style anonymous payments for citizens, and income-transparency for taxability.

- Participants: Florian Dold, Gabor Toth, Hans Grothoff, Jeffrey Paul Burdges and Marcello Stanisci
- Partner: The GNU Project
- Contact: Hans Grothoff
- URL: http://taler.net/

## 6.5. HyLeak

*Hybrid Analysis Tool for Information Leakage* KEYWORD: Information leakage FUNCTIONAL DESCRIPTION: HyLeak is an evolution of the QUAIL tool, also developed by the TAMIS team. HyLeak divides the input program into (terminal) components and decides for each of them whether to analyze it using precise or statistical analysis, by applying heuristics that evaluate the analysis cost of each component. Then, HyLeak composes the analysis results of all components into an approximate joint probability distribution of the secret and observable variables in the program. Finally, the tool estimates the Shannon leakage and its confidence interval.

- Partner: AIST Tsukuba
- Contact: Fabrizio Biondi

# 6.6. SimFI

#### Tool for Simulation Fault injection

**KEYWORDS:** Fault injection - Fault-tolerance

FUNCTIONAL DESCRIPTION: Fault injections are used to test the robust and security of systems. We have developed SimFI, a tool that can be used to simulate fault injection attacks against binary files. SimFI is lightweight utility designed to be integrated into larger environments as part of robustness testing and fault injection vulnerability detection.

- Contact: Nisrine Jafri
- URL: https://github.com/nisrine/Fault-Injection-Tool

## 6.7. DaD

#### Data-aware Defense

KEYWORD: Ransomware

FUNCTIONAL DESCRIPTION: DaD is a ransomware countermeasure based on a file system minifilter driver. It is a proof of concept and in its present condition cannot be used as a replacement of the existing antivirus solutions. DaD detects randomness of the data by monitoring the write operations on the file system. We monitor all the userland threads, and also the whole file system (i.e., not restricted to Documents). It blocks the threads that exceed a specific threshold. The malicious thread is not killed, we only block its next I/O operations.

• Contact: Aurélien Palisse

## 6.8. MASSE

#### Modular Automated Syntactic Signature Extraction

KEYWORDS: Malware - Syntactic analysis

FUNCTIONAL DESCRIPTION: The Modular Automated Syntactic Signature Extraction (MASSE) architecture is a new integrated open source client-server architecture for syntactic malware detection and analysis based on the YARA, developed with Teclib'. MASSE includes highly effective automated syntactic malware detection rule generation for the clients based on a server-side modular malware detection system. Multiple techniques are used to make MASSE effective at detecting malware while keeping it from disrupting users and hindering reverse-engineering of its malware analysis by malware creators. MASSE integrates YARA in a distributed system able to detect malware on endpoint systems using YARA, analyze malware with multiple analysis techniques, automatically generate syntactic malware detection rules, and deploy the new rules to the endpoints. The MASSE architecture is freely available to companies and institutions as a complete, modular, self-maintained antivirus solution. Using MASSE, a security department can immediately update the rule database of the whole company, stopping an infection on its tracks and preventing future ones.

• Contact: Axel Legay

## **6.9. Behavioral Malware Analysis**

KEYWORDS: Artificial intelligence - Malware - Automatic Learning - Concolic Execution

FUNCTIONAL DESCRIPTION: Our approach is based on artificial intelligence. We extract graphs from programs, that represent their behaviors. Such graphs are called system call dependency graphs (SCDGs). Our software learns to distinguish malware from cleanware on a large set of malwares and cleanwares. Whenever we want to analyze a new program, we extract its graphs and use the result of the training to decide whether the new program to analyze is a malware.

- Partner: Cisco
- Contact: Axel Legay
- URL: https://team.inria.fr/tamis/

# 6.10. VITRAIL - Visualisation Tool

### Real-Time, Advanced, Immersive Visualization of Software / Visualizer

KEYWORD: Visualization of software

SCIENTIFIC DESCRIPTION: It is difficult for developers to explore and understand the source code of large programs, for example in objet-oriented languages programs featuring thousands of classes. Visualization methods based on daily life metaphors have thus been proposed. The VITRAIL Visualization tool (or VITRAIL Vizualizer) makes it possible to display, visualize and explore Java programs in a metaphorical way, using the city metaphor. An execution trace of the Java (byte)code provided by VITRAIL JBInstrace tool, is provided as input to VITRAIL Visualizer which displays a city-like metaphorical world showing the static structure of the code as well as some dynamic elements (calls).

FUNCTIONAL DESCRIPTION: This program makes it possible to displays, visualizes and explores Java programs in a metaphorical way (using the city metaphore). Useful for complex application developers/architects.

RELEASE FUNCTIONAL DESCRIPTION: Early release

- Participants: Damien Bodenes, Olivier Demengeon and Olivier Zendra
- Contact: Olivier Zendra
- URL: http://vitrail.loria.fr

## 6.11. VITRAIL 6 JBInsTrace

*Real-Time, Advanced, Immersive Visualization of Software / Java Bytecode Instrumenter and Tracer* KEYWORDS: Execution trace - Profiling - Instrumentation - Bytecode - Java - Basic block

SCIENTIFIC DESCRIPTION: VITRAIL JBInsTrace is a program to instrument Java bytecode to trace its execution. The trace contains both static and dynamic information (calls). It is produced by intercepting the JVM class loader and replacing it by ours. Thus Java bytecode file are not modified, since instrumentation is performed on the fly, in memory. This makes it possible to instrument the whole program code, including libraries. Java source code is not needed. The trace which is then fed into our program VITRAIL Visualizer is an XML-like file.

FUNCTIONAL DESCRIPTION: VITRAIL JBInsTrace is a program to instrument Java bytecode files to trace their execution. The trace is then fed into our VITRAIL Visualizer tool.

- Participants: Olivier Zendra and Pierre Caserta
- Contact: Olivier Zendra
- URL: http://vitrail.loria.fr

## 6.12. Platforms

## 6.12.1. Malware'o'Matic

This LHS platform is dedicated to the collect, the categorization and the analyze of malware. We are currently interested in a specific kind of malware the ransomware. The platform grabs periodically samples of public data bases, executes the ransomware without virtualization on a victim PC and evaluate the implemented detection mechanisms. Once a ransomware has been executed the image of the OS is automatically restored and a new sample is evaluated. The platform is fully automatic and target Windows platforms (seven, W10) in both 32 bits and 64 bits versions. More recent developments can be seen in the LHS Activity Report.

## 6.12.2. Faustine

This LHS platform is dedicated to the EM fault injection experiments. It is composed of a motion table (XY), a pulse generator, an amplifier and a control PC. It injects EM pulses in a controlled way on a targeted device using an EM probe. It controls with a high precision the timing and the edges of the pulse. A recent development consists in adding a FPGA board to control the trigger in a more convenient and precise way. Then, the pulse can be triggered while a specific information is sent to the board under attack. More recent developments can be seen in the LHS Activity Report.

# 7. New Results

# 7.1. Results for Axis 1: Vulnerability analysis

### 7.1.1. Statistical Model Checking of LLVM Code

Participants: Axel Legay, Louis-Marie Traonouez.

We have extended PLASMA Lab statistical model-checker with a new plugin that allows to simulate LLVM bitcode. The plugin is based on an external simulator LODIN. This simulator implements a probabilistic semantics for a LLVM program. At its core the semantics consist of the LLVM program given as a labelled transition system. The labels are function calls to an environment that implements functions outside the LLVM core language. The environment is also responsible for assigning probabilities to individual transitions

By interfacing the LODIN simulator with PLASMA Lab we can apply all the statistical model-checking algorithms provided by PLASMA Lab, including rare events verification algorithms like importance splitting. We have applied LODIN and PLASMA Lab to several case studies, including the analysis of some security vulnerability, like the PTrace privilege escalation attack that could be performed on earlier versions of the Linux Kernel. This work has been submitted to a conference this year [61], and is currently under review.

[61] We present our work in providing Statistical Model Checking for programs in LLVM bitcode. As part of this work we develop a semantics for programs that separates the program itself from its environment. The program interact with the environment through function calls. The environment is furthermore allowed to perform actions that alter the state of the C-program-useful for mimicking an interrupt system. On top of this semantics we build a probabilistic semantics and present an algorithm for simulating traces under that semantics. This paper also includes the development of the new tool component Lodin that provides a statistical model checking infrastructure for LLVM programs. The tool currently implement standard Monte Carlo algorithms and a simulator component to manually inspect the behaviour of programs. The simulator also proves useful in one of our other main contributions; namely producing the first tool capable of doing importance splitting on LLVM code. Importance splitting is implemented by integrating Lodin with the existing statistical model checking tool Plasma-Lab.

#### 7.1.2. Verification of IKEv2 protocol

Participants: Axel Legay, Tristan Ninet, Louis-Marie Traonouez, Olivier Zendra.

The IKEv2 (Internet Key Exchange version 2) protocol is the authenticated key-exchange protocol used to set up secure communications in an IPsec (Internet Protocol security) architecture. It guarantees security properties like mutual-authentication and secrecy of the exchanged key. To obtain an IKEv2 implementation as secure as possible, we use model checking to verify the properties on the protocol specification, and smart fuzzing to test the implementation, and try to detect implementation flaws like buffer overflows or memory leaks.

Two weaknesses had previously been found in the specification, but were harmless. We showed that the first weakness does not actually exist. We demonstrated that the second weakness is not harmless, and we designed a Denial-of-Service attack that exploits it, the deviation attack. As a counter-measure, we propose a modification of IKEv2, and use model checking to prove that the modified version is secure.

This work is being prepared for responsive disclosure and publication.

#### 7.1.3. High-Level Frameworks for Scheduling Systems

Participants: Mounir Chadli, Axel Legay, Louis-Marie Traonouez.

Formal model-based techniques are more and more used for the specification and verification of scheduling systems. These techniques allow to consider complex scheduling policies beyond the scope of classical analytical techniques. For instance, hierarchical scheduling systems (HSS) integrates a number of components into a single system running on one execution platform. Hierarchical scheduling systems have been gaining more attention by automotive and aircraft manufacturers because they are practical in minimizing the cost and energy of operating applications. Model-based techniques can also be used to solve new problems like energy optimization or runtime monitoring. However, one limitation of formal model-based approaches is that they require high technical knowledge about the formalims and tools used to design models and write properties.

In a previous work [62], we have presented a model-based framework for the verification of HSS. It is based on a stochastic extension of timed automata and statistical model checking with the tool UPPAAL. We have also developed a graphical high-level language to represent complex hierarchical scheduling systems. To bridge the gap between the formalisms, we exploit Cinco, a generator for domain specific modeling tools to generate an interface between this language and the one of UPPAAL. Cinco allows to specify the features of a graphical interface in a compact meta-model language. This is a flexible approach that could be extended to any formal model of scheduling problem.

We have extended the previous work in journal paper [55] published this year, where we provide another highlevel framework for the verification of energy-aware scheduling systems. We also present two new analysis techniques. One that performs runtime monitoring in order to detect alarming change in the scheduling system, and one that performs energy optimization.

[55] Over the years, schedulability of Cyber-Physical Systems (CPS) has mainly been performed by analytical methods. These techniques are known to be effective but limited to a few classes of scheduling policies. In a series of recent work, we have shown that schedulability analysis of CPS could be performed with a model-based approach and extensions of verification tools such as UPPAAL. One of our main contributions has been to show that such models are flexible enough to embed various types of scheduling policies, which goes beyond those in the scope of analytical tools.

However, the specification of scheduling problems with model-based approaches requires a substantial modeling effort, and a deep understanding of the techniques employed in order to understand their results. In this paper we propose simplicity-driven high-level specification and verification frameworks for various scheduling problems. These frameworks consist of graphical and user-friendly languages for describing scheduling problems. The high-level specifications are then automatically translated to formal models, and results are transformed back into the comprehensible model view. To construct these frameworks we exploit a meta-modeling approach based on the tool generator Cinco.

Additionally we propose in this paper two new techniques for scheduling analysis. The first performs runtime monitoring using the CUSUM algorithm to detect alarming change in the system. The second performs optimization using efficient statistical techniques. We illustrate our frameworks and techniques on two case studies.

### 7.1.4. Side-channel Analysis of Cryptographic Substitution Boxes

Participants: Axel Legay, Annelie Heuser.

With the advent of the Internet of Things, we are surrounded with smart objects (aka things) that have the ability to communicate with each other and with centralized resources. The two most common and widely noticed artefacts are RFID and Wireless Sensor Networks which are used in supply-chain management, logistics, home automation, surveillance, traffic control, medical monitoring, and many more. Most of these applications have the need for cryptographic secure components which inspired research on cryptographic

algorithms for constrained devices. Accordingly, lightweight cryptography has been an active research area over the last 10 years. A number of innovative ciphers have been proposed in order to optimize various performance criteria and have been subject to many comparisons. Lately, the resistance against side-channel attacks has been considered as an additional decision factor.

Side-channel attacks analyze physical leakage that is unintentionally emitted during cryptographic operations in a device (e.g., power consumption, electromagnetic emanation). This side-channel leakage is statistically dependent on intermediate processed values involving the secret key, which makes it possible to retrieve the secret from the measured data.

Side-channel analysis (SCA) for lightweight ciphers is of particular interest not only because of the apparent lack of research so far, but also because of the interesting properties of substitution boxes (S-boxes). Since the nonlinearity property for S-boxes usually used in lightweight ciphers (i.e.,  $4 \times 4$ ) can be maximally equal to 4, the difference between the input and the output of an S-box is much smaller than for instance for AES. Therefore, one could conclude that from that aspect, SCA for lightweight ciphers must be more difficult. However, the number of possible classes (e.g., Hamming weight (HW) or key classes) is significantly lower, which may indicate that SCA must be easier than for standard ciphers. Besides the difference in the number of classes and consequently probabilities of correct classification, there is also a huge time and space complexity advantage (for the attacker) when dealing with lightweight ciphers.

In [65], [64] we give a detailed study of lightweight ciphers in terms of side-channel resistance, in particular for software implementations. As a point of exploitation we concentrate on the non-linear operation (S-box) during the first round. Our comparison includes SPN ciphers with 4-bit S-boxes such as KLEIN, PRESENT, PRIDE, RECTANGLE, Mysterion as well as ciphers with 8-bit S-boxes: AES, Zorro, Robin. Furthermore, using simulated data for various signal-to-noise ratios (SNR) we present empirical results for Correlation Power Analysis (CPA) and discuss the difference between attacking 4-bit and 8-bit S-boxes.

An extension of this work is given in [10]. We investigate whether side-channel analysis is easier for lightweight ciphers than e.g. for AES. We cover both profiled and non-profiled techniques where we are interested in recovering secret (round)keys or intermediate states. In the case of non-profiled attacks, we evaluate a number of S-boxes appearing in lightweight ciphers using the confusion coefficient and empirical simulations.

First, we investigate in the scenario where the attacker targets the first round and thus exploits the S-box computation. We observe that the 8-bit S-boxes from AES, Zorro, and Robin perform similarly, whereas for 4-bit S-boxes we have a clear ranking, with the S-box of Piccolo being the weakest to attack and the S-box of KLEIN and Midori (1) the hardest. Interestingly, when considering the last round and thus the inverse S-box operation the ranking changes such that Mysterion is the weakest and PRESENT/LED is the most side-channel resistant cipher from the ones investigated. Moreover, we could observe that attacking the last round is equal or less efficient for all considered ciphers. Finally, we use the information gained from both rounds together, where this approach is of interest when the cipher does not use round keys from a key scheduling algorithm but rather uses the same (or a straightforward computable) key in each round. LED fulfils this requirement. For a reasonable low SNR, to reach a success rate of 0.9 an attack on both rounds only requires 100 traces, whereas an attack using the first round requires 200 traces and on the last 400 traces. This example highlights the important role the confusion coefficient (relationship between predicted intermediate states under a leakage model from different key hypotheses), and that not only the SNR (even if low) is a key factor influencing the success rate. Additionally, our result show that we cannot conclude that the 4-bit S-boxes are generally significantly less resistant than the investigated 8-bit S-boxes. In particular, when considering inverse S-boxes we showed that 4-bit S-boxes may be more resistant.

For profiled attacks, we analyze several machine learning techniques to recover 4-bit and 8-bit intermediate states. Our results show that attacking 4-bit is somewhat easier than attacking 8-bit, with the difference mainly stemming from the varying number of classes in one or the other scenario. Still, that difference is not so apparent as one could imagine. Since we work with only a single feature and yet obtain a good accuracy in a number of test scenarios, we are confident (as our experiments also confirm) that adding more features will render classification algorithms even more powerful, which will result in an even higher accuracy. Finally, we

did not consider any countermeasures for the considered lightweight algorithms, since the capacity for adding countermeasures is highly dependent on the environment (which we assume to be much more constrained than in the case of AES). However, our results show that a smart selection of S-boxes results in an inherent resilience (especially for 4-bit S-boxes). Moreover, we show that in case of highly restricted devices, in which countermeasures on the whole cipher are not practically feasible, a designer may choose to only protect the weakest round (first round) in the cipher to increase the side-channel resistant until a certain limit.

Our work in [23] concentrates on how to improve SCA resilience of ciphers without imposing any extra cost. This is possible by considering the inherent resilience of ciphers. We particularly concentrate on block ciphers which utilize S-boxes and therefore study the resilience of S-boxes against side-channel attacks. When discussing how to improve side-channel resilience of a cipher, an obvious direction is to use various masking or hiding countermeasures. However, such schemes come with a cost, e.g. an increase in the area and/or reduction of the speed. When considering lightweight cryptography and various constrained environments, the situation becomes even more difficult due to numerous implementation restrictions. However, some options are possible like using S-boxes that are easier to mask or (more on a fundamental level), using S-boxes that possess higher inherent side-channel resilience. In [23] we investigate what properties should an S-box possess in order to be more resilient against side-channel attacks. Moreover, we find certain connections between those properties and cryptographic properties like nonlinearity and differential uniformity. Finally, to strengthen our theoretical findings, we give an extensive experimental validation of our results.

- [64] Side-channel Analysis of Lightweight Ciphers: Current Status and Future Directions
- [65] Side-channel Analysis of Lightweight Ciphers: Does Lightweight Equal Easy?
- [10] Lightweight Ciphers and their Side-channel Resilience.
- [23] Trade-Offs for S-Boxes: Cryptographic Properties and Side-Channel Resilience
- [24] Do we need a holistic approach for the design of secure IoT systems? hal-01628683

#### 7.1.5. New Advances on Side-channel Distinguishers

Participants: Axel Legay, Annelie Heuser.

[16] Template Attack vs Bayes Classifier

Side-channel attacks represent one of the most powerful category of attacks on cryptographic devices with profiled attacks in a prominent place as the most powerful among them. Indeed, for instance, template attack is a well-known real-world attack that is also the most powerful attack from the information theoretic perspective. On the other hand, machine learning techniques have proven their quality in a numerous applications where one is definitely side-channel analysis. As one could expect, most of the research concerning supervised machine learn- ing and side-channel analysis concentrated on more powerful machine learning techniques. Although valid from the practical perspective, such attacks often remain lacking from the more theoretical side. In this paper, we investigate several Bayes classifiers, which present simple supervised techniques that have significant similarities with the template attack. More specifically, our analysis aims to investigate what is the influence of the feature (in)dependence in datasets with different amount of noise and to offer further insight into the efficiency of machine learning for side-channel analysis.

[46] Side-channel analysis and machine learning: A practical perspective The field of side-channel analysis has made significant progress over time. Analyses are now used in practice in design companies as well as in test laboratories, and the security of products against side-channel attacks has significantly improved. However, there are still some remaining issues to be solved for analyses to be more effective. Side-channel analysis ac- tually consists of two steps, commonly referred to as identification and exploitation. The identification consists of understanding the leakage in order to set up a relevant attack. On the other hand, the exploitation consists of using the identified leakages to extract the secret key. In scenarios where the model is poorly known, it can be approximated in a profiling phase. There, machine learning techniques are gaining value. In this paper, we conduct extensive analysis of several machine learning techniques, showing the importance of proper

parameter tuning and training. In contrast to what is perceived as common knowledge in unrestricted scenarios, we show that some machine learning techniques can significantly outperform template attack when properly used. We therefore stress that the traditional worst case security assessment of cryptographic implementations that includes mainly template attacks might not be accurate enough. Besides that, we present a new measure called the Data Confusion Factor that can be used to assess how well machine learning techniques will perform on a certain dataset.

[30] Codes for Side-Channel Attacks and Protections

This article revisits side-channel analysis from the standpoint of coding theory. On the one hand, the attacker is shown to apply an optimal decoding algorithm in order to recover the secret key from the analysis of the side-channel. On the other hand, the side-channel protections are presented as a coding problem where the information is mixed with randomness to weaken as much as possible the sensitive information leaked into the side-channel. Therefore, the field of side-channel analysis is viewed as a struggle between a coder and a decoder. In this paper, we focus on the main results obtained through this analysis. In terms of attacks, we discuss optimal strategy in various practical contexts, such as type of noise, dimensionality of the leakage and of the model, etc. Regarding countermeasures, we give a formal analysis of some masking schemes.

[38] Climbing Down the Hierarchy: Hierarchical Classification for Machine Learning Side-Channel Attacks

Machine learning techniques represent a powerful paradigm in side-channel analysis, but they come with a price. Selecting the appropriate algorithm as well as the parameters can sometimes be a difficult task. Nevertheless, the results obtained usually justify such an effort. However, a large part of those results use simplification of the data relation and in fact do not consider allthe available information. In this paper, we analyze the hierarchical relation between the data and propose a novel hierarchical classification approach for side-channel analysis. With this technique, we are able to introduce two new attacks for machine learning side-channel analysis: Hierarchical attack and Structured attack. Our results show that both attacks can outperform machine learning techniques using the traditional approach as well as the template attack regarding accuracy. To support our claims, we give extensive experimental results and discuss the necessary conditions to conduct such attacks.

[14] Stochastic Collision Attack

On the one hand, collision attacks have been introduced in the context of side-channel analysis for attackers who exploit repeated code with the same data without having any knowledge of the leakage model. On the other hand, stochastic attacks have been introduced to recover leakage models of internally processed intermediate secret variables. Both techniques have shown advantages and intrinsic limitations. Most collision attacks, for instance, fail in exploiting all the leakages (e.g., only a subset of matching samples are analyzed), whereas stochastic attacks cannot involve linear regression with the full basis (while the latter basis is the most informative one). In this paper, we present an innovative attacking approach, which combines the flavors of stochastic and collision attacks. Importantly, our attack is derived from the optimal distinguisher, which maximizes the success rate when the model is known. Notably, we develop an original closed-form expression, which shows many benefits by using the full algebraic description of the leakage model. Using simulated data, we show in the unprotected case that, for low noise, the stochastic collision attack is superior to the state of the art, whereas asymptotically and thus, for higher noise, it becomes equivalent to the correlation-enhanced collision attack. Our so-called stochastic collision attack is extended to the scenario where the implementation is protected by masking. In this case, our new stochastic collision attack is more efficient in all scenarios and, remarkably, tends to the optimal distinguisher. We confirm the practicability of the stochastic collision attack thanks to experiments against a public data set (DPA contest v4). Furthermore, we derive the stochastic collision attack in case of zero-offset leakage that occurs in protected hardware implementations and use simulated data for comparison. Eventually, we underline the capability of the new distinguisher to improve its efficiency when the attack multiplicity increases.

#### [15] Optimal side-channel attacks for multivariate leakages and multiple models

Side-channel attacks allow to extract secret keys from embedded systems like smartcards or smartphones. In practice, the side-channel signal is measured as a trace consisting of several samples. Also, several sensitive bits are manipulated in parallel, each leaking differently. Therefore, the informed attacker needs to devise side-channel distinguishers that can handle both multivariate leakages and multiple models. In the state of the art, these two issues have two independent solutions: on the one hand, dimensionality reduction can cope with multivariate leakage; on the other hand, online stochastic approach can cope with multiple models. In this paper, we combine both solutions to derive closed-form expressions of the resulting optimal distinguisher in terms of matrix operations, in all situations where the model can be either profiled offline or regressed online. Optimality here means that the success rate is maximized for a given number of traces. We recover known results for uni- and bivariate models (including correlation power analysis) and investigate novel distinguishers for multiple models with more than two parameters. In addition, following ideas from the AsiaCrypt?2013 paper ?Behind the Scene of Side-Channel Attacks,? we provide fast computation algorithms in which the traces are accumulated prior to computing the distinguisher values.

#### [39] Stochastic Side-Channel Leakage Analysis via Orthonormal Decomposition

Side-channel attacks of maximal efficiency require an accurate knowledge of the leakage function. Template attacks have been introduced by Chari et al. at CHES 2002 to estimate the leakage function using available training data. Schindler et al. noticed at CHES 2005 that the complexity of profiling could be alleviated if the evaluator has some prior knowledge on the leakage function. The initial idea of Schindler is that an engineer can model the leakage from the structure of the circuit. However, for some thin CMOS technologies or some advanced countermeasures, the engineer intuition might not be sufficient. Therefore, inferring the leakage function based on profiling is still important. In the state-of-the-art, though, the profiling stage is conducted based on a linear regression in a non-orthonormal basis. This does not allow for an easy interpretation because the components are not independent. In this paper, we present a method to characterize the leakage based on a Walsh-Hadamard orthonormal basis with staggered degrees, which allows for direct interpretations in terms of bits interactions. A straightforward application is the characterization of a class of devices in order to understand their leakage structure. Such information is precious for designers and also for evaluators, who can devise attack bases relevantly.

#### [17] On the optimality and practicability of mutual information analysis in some scenarios

The best possible side-channel attack maximizes the success rate and would correspond to a maximum likelihood (ML) distinguisher if the leakage probabilities were totally known or accurately estimated in a profiling phase. When profiling is unavailable, however, it is not clear whether Mutual Information Analysis (MIA), Correlation Power Analysis (CPA), or Linear Regression Analysis (LRA) would be the most successful in a given scenario. In this paper, we show that MIA coincides with the maximum likelihood expression when leakage probabilities are replaced by online estimated probabilities. Moreover, we show that the calculation of MIA is lighter that the computation of the maximum likelihood. We then exhibit two case-studies where MIA outperforms CPA. One case is when the leakage model is known but the noise is not Gaussian. The second case is when the leakage model is partially unknown and the noise is Gaussian. In the latter scenario MIA is more efficient than LRA of any order.

#### [59] On the Relevance of Feature Selection for Profiled Side-channel Attacks

In the process of profiled side-channel analysis there is a number of steps one needs to make. One important step that is often conducted without a proper attention is selection of the points of interest (features) within the side-channel measurement trace. Most of the related work start with an assumption that the features are selected and various attacks are then considered and compared to find the best approach. In this paper, we concentrate on the feature selection step and show that if a proper selection is done, most of the attack techniques offer satisfactory results. We investigate how more advanced feature selection techniques stemming from the machine learning domain can be used to improve the side-channel attack efficiency. Our results show that the so-called Hybrid feature selection methods result in the best classification accuracy over a wide range of test scenarios and number of features selected.

[60] Profiled SCA with a New Twist: Semi-supervised Learning

Profiled side-channel attacks represent the most powerful category of side-channel attacks. In this context, the attacker gains ac- cess of a profiling device to build a precise model which is used to attack another device in the attacking phase. Mostly, it is assumed that the attacker has unlimited capabilities in the profiling phase, whereas the attacking phase is very restricted. We step away from this assumption and consider an attacker who is restricted in the profiling phase, while the attacking phase is less limited as in the traditional view. Clearly, in general, the attacker is not hindered to exchange any available knowledge between the profiling and attacking phase. Accordingly, we propose the concept of semi-supervised learning to side-channel analysis, in which the attacker uses the small amount of labeled measurements from the profiling phase as well as the unlabeled measurements from the attacking phase to build a more reliable model. Our results show that semi-supervised learning is beneficial in many scenarios and of particular interest when using template attack and its pooled version as side-channel attack techniques. Besides stating our results in varying scenarios, we discuss more general conclusions on semi-supervised learning for SCA that should help to transfer our observations to other settings in SCA.

#### 7.1.6. Side-channel analysis on post-quantum cryptography

Participants: Axel Legay, Annelie Heuser, Tania Richmond, Martin Moreau.

In recent years, there has been a substantial amount of research on quantum computers ? machines that exploit quantum mechanical phenomena to solve mathematical problems that are difficult or intractable for conventional computers. If large-scale quantum computers are ever built, they will be able to break many of the public-key cryptosystems currently in use. This would seriously compromise the confidentiality and integrity of digital communications on the Internet and elsewhere. The goal of post-quantum cryptography (also called quantum-resistant cryptography) is to develop cryptographic systems that are secure against both quantum and classical computers, and can interoperate with existing communications protocols and networks. At present, there are several post-quantum cryptosystems that have been proposed: lattice-based, code-based, multivariate cryptosystems, hash-based signatures, and others. However, for most of these proposals, further research is needed in order to gain more confidence in their security and to improve their performance. Our interest lies in particular on the side-channel analysis and resistance of these post-quantum schemes. We first focus on code-based cryptography and then extend our analysis to find common vulnerabilities between different families of post-quantum crypto systems.

### 7.1.7. Binary Code Analysis: Formal Methods for Fault Injection Vulnerability Detection

Participants: Axel Legay, Thomas Given-Wilson, Annelie Heuser, Nisrine Jafri, Jean-Louis Lanet.

Formal methods such as model checking provide a powerful tool for checking the behaviour of a system. By checking the properties that define correct system behaviour, a system can be determined to be correct (or not).

Increasingly fault injection is being used as both a method to attack a system by a malicious attacker, and to evaluate the dependability of the system. By finding fault injection vulnerabilities in a system, the resistance to attacks or faults can be detected and subsequently addressed.

A process is presented that allows for the automated simulation of fault injections. This process proceeds by taking the executable binary for the system to be tested, and validating the properties that represent correct system behaviour using model checking. A fault is then injected into the executable binary to produce a mutant binary, and the mutant binary is model checked also. A different result to the validation of the executable binary in the checking of the mutant binary indicates a fault injection vulnerability.

This process has been automated with existing tools, allowing for easy checking of many different fault injection attacks and detection of fault injection vulnerabilities. This allows for the detection of fault injection vulnerabilities to be fully automated, and broad coverage of the system to be formally shown.

The work is implemented in the SimFi tool.

[56] (J; submitted) Fault injection has increasingly been used both to attack software applications, and to test system robustness. Detecting fault injection vulnerabilities has been approached with a variety of different but limited methods. This paper proposes an extension of a recently published general model checking based process to detect fault injection vulnerabilities in binaries. This new extension makes the general process scalable to real-world implementations which is demonstrated by detecting vulnerabilities in different cryptographic implementations.

#### 7.1.8. Security at the hardware and software boundaries

Participants: Axel Legay, Jean-Louis Lanet, Ronan Lashermes, Kevin Bukasa, Hélène Le Bouder.

#### 7.1.8.1. Side-channel attacks (SCA)

SCA exploit the reification of a computation through its physical dimensions (current consumption, EM emission, etc.). Focusing on Electromagnetic Analyses (EMA), such analyses have mostly been considered on low-end devices: smartcards and micro-controllers. In the wake of recent works, we analyze the effects of a modern micro architecture [31] on the efficiency of EMA (here Correlation Power Analysis and template attacks). We show that despite the difficulty to synchronize the measurements, the speed of the targeted core and the activity of other cores on the same chip can still be accommodated. Finally, we confirm that enabling the secure mode of TrustZone (a hardware-assisted software countermeasure) has no effect whatsoever on the EMA efficiency. Therefore, critical applications in TrustZone are not more secure than in the normal world with respect to EMA, in accordance with the fact that it is not a countermeasure against physical attacks. We hint that such techniques may be more common in the future to overcome the true difficulty with high-end devices: dealing with time precision (problem even worse with an OS or a virtual machine). Here again TrustZone or the activity of other cores have no incidence. But with these attacks, managing the big amount of data generated by our measures may prove to be the limiting factor, requiring better computing resources.

We investigate the way the compiler works and new attack paths have been discovered. In particular we demonstrated experimentally on an ARM7m the possibility to execute arbitrary code, generate buffer overflow even in presence of compiler assisted canary and ROP attacks. This raises a new challenge: any code fragment of an embedded program is sensitive to a fault attack. Thus an attacker increases the success rate of its attack while targeting a non sensitive part of the program for the injection. Then it becomes easy to extract security materials from the device. Then, the verification of the absence of a potential vulnerability must be checked on the whole program and not only on the cryptographic primitives. Thus the prevention analysis that was possible thanks to formal methods becomes unreachable with these new attack paths [40].

#### 7.1.8.2. SCA based fuzzer

One of the main challenges during the development of system is to give a proof of evidence that its functionalities are correctly implemented and that no vulnerability remains. This objective is mostly achieved via testing techniques, which include software testing to check whether a system meets its functionalities, or security testing to express what should not happen. For the latter case, fuzzing is considered as first class citizen. It consists in exercising the system with (randomly) generated and eventually modified inputs in order to test its resistance. While fuzzing is definitively the fastest and the easiest way for testing applications, it suffers from severe limitations. Indeed, the precision of the model used for input generation: a random and/or simple model cannot reach all states and significant values. Moreover, a higher model precision can result in a combinatorial explosion of test cases.

We suggest a new approach [11] whose main ingredient is to combine timing attacks with fuzzing techniques. This new approach, allows not only reducing the test space explosion, but also to simplify the fuzzing process configuration. This new testing scenario is based on observing several executions of the system and by freezing some of its parameters in order to establish a partial order on their timing evaluation. The root of our technique is to exploit timing information to classify the input data into sub-domains according to the behavior observed for specific values of the parameters. Our approach is able to discover hidden unspecified commands that may trigger computations in the tested software. Due to the specific nature of the application (the domain of the parameters is the byte) and its programming model we can also retrieve the control flow graph of

the application. The limits of the approach have been identified, and it has been tested on two applications. Validation via a coverage tool has been established.

## 7.1.9. System Vulnerability Analysis

Participants: Jean-Louis Lanet, Abdelhal Mesbah, Razika Lounas, Chaharezd Yayaoui.

We present in this section our effort to detect and correct some misbehaviors encountered with some firmware. We start with an attack on a secure device, such that we are able to reverse a code while the ISA is unknown and the code itself is not available. Then, we propose a formal specification of the update process of a firmware which provides the guarantee that the updated program respects the semantics of the language. In a last aspect, we try to predict the ability of a program to be attacked thanks to a Machine Learning algorithm. We demonstrated in section 7.1.8 that a state exploration is useless until the whole program is examined, we demonstrated here that approximative solutions can deal with real live programs with an affordable response time.

#### 7.1.9.1. Reverse engineering

We believe that an adversary can gain access to different assets of the system using a black box approach. This implies of course the absence of the source code, but also sometime the absence of the binary code (romized within the soc or micro-controller, no update mechanism, no jtag, no memory extraction, no read function, and so on). In that case, the first step consists in extracting the binary code from the system. The attacker is just allowed to load data. He has then to infer enough information on the system internals and then he should be able to gain access to the native layers. In [43], we demonstrate the advantage of a graphical representation of the data in the memory can help the reverse process thanks to the abstraction provided. Our graphical tool links all the objects with a relationship based on the presence of a pointer.

In a Java based secure element, a Java application is considered as data executed by the executed program (the virtual machine) by the native processor. We introduce a first weakness in the program that allows to read an instance as an array which violate the Java type system. This weakness allows us to dump a short part of the memory which contains the meta data on a set of arrays. Thanks to this information, we generate a mimicry attack by forging pointer illegally [41]. In turns, it open the possibility to read large part of the memory as element of a forged array. Then we succeed in characterizing the memory management algorithm [12]. At the end, we transform the initial problem of finding a vulnerability in the code of a device in a black box approach to a white box problem after de-assembling the binary code.

In another work [44], we studied the byte code verification process towards an unchecked code. We found that this verification is not complete and can be bypassed. The verifier checks the semantics of the Java Card byte code. This process is split in two parts. First, the verifier loads the methods' byte code and checks the package content. For the method segment, it checks that the control flow remain inside the methods, the jump destinations are correct and so on. Secondly, for each entry point and only for these, it controls the semantics and the type correctness of the code. This step is not performed for unreachable code, while the specification states that no unreachable code should remain in the file. However, during our analysis we discovered that the verifier does some verification on the semantics of the unreachable code. Then, thanks to a fault attack (the return byte code is noped) we diverted the control flow into this unchecked area were we stored our ill-typed code leading to the execution of an aggressive shell code which in turn dumped the native layers of the card giving access to the secret key material in plain text.

#### 7.1.9.2. Safe system update mechanism

Dynamic Software Updating (DSU) consists in updating running programs on the fly without any downtime. This feature is interesting in critical applications that must run continuously. Because updates may lead to security breaches, the question of their correctness is raised. Formal methods are a rigorous means to ensure the correctness required by applications using DSU. We propose [13] a formal verification of correctness of DSU in a Java-based embedded system. Our approach is based on three steps. First, a formal interpretation of the semantics of update operations to ensure type safety of the update. Secondly, we rely on a functional representation of byte code, the predicate transformation calculus, and a functional model of the update

mechanism to ensure the behavioral correctness of the updated programs. It is based on the use of Hoare predicate transformation to derive a specification of an updated byte code. In the last step, we use the functional representation to model the safe update point detection mechanism. This mechanism guarantees that none of the updated method active methods are active. This property is called activeness safety. We propose a functional specification that allows to derive proof obligations that guarantee the safety of the mechanism.

#### 7.1.9.3. Prediction of system divergence

Fault attack represents one of the serious threats against embedded system security. The result of the fault injection could lead to a mutation of the code in such a way that it becomes hostile or execute a unwanted sequence of code as we demonstrated in 7.1.8. Any successful attack may reveal a secret information stored in the card or grant an undesired authorization. We propose a methodology [5] to recognize, during the development step, the sensitive patterns to the fault attack. It is based on the concepts from text categorization and machine learning. In fact, in this method we represented the patterns using opcodes n-grams as features and we evaluated different machine learning classifiers.

In the first experiment, we evaluated all the combination of n-gram size (for n=2, n=3 and n=4), number of features using GR method to select 100, 200, ..., 500 and 1000 relevant n-grams, n-gram weighting (Term Frequency (TF), Term Frequency Inverse Document Frequency (TFIDF) and binary representations), and five classification algorithms (Naive Bayes network (NB), Decision Tree (DT), Support Vector Machine (SVM), and the boosted version of these two lasts (BDT and BSVM)) to determine the best setting. We used accuracy measure to evaluate performance of the classifiers. In addition to accuracy, we used F1, TP rate and FP rate measures to evaluate how the algorithms classified the dangerous patterns. In the first experiment, we noted that 2-gram outperformed others. Nearly 2-gram, TFIDF, 1000 features with boosted algorithm outperformed the other settings. The F1 results have shown that the classifiers are more accurate at classifying examples of the class of non dangerous pattern compared to other classes. We suggest that this might be due to the imbalance of our data set. In the second experiment, we investigated the imbalance problem. We applied SMOTE and NCR resampling techniques to overcome this class imbalance problem. We found that the outperforming setting in the resampled data set was St<sub>270</sub> also with BSVM classifier. Resampled data set improves accuracy of the smallest class and keeps the accuracy of other classes.

The experimental results indicated that the resampling techniques improved the accuracy of the classifiers. In addition, our proposed method reduces the execution time of sensitive patterns classification in comparison to the mutant generator tool micro seconds instead of hours.

## 7.2. Results for Axis 2: Malware analysis

The detection of malicious programs is a fundamental step to be able to guarantee system security. Programs that exhibit malicious behavior, or *malware*, are commonly used in all sort of cyberattacks. They can be used to gain remote access on a system, spy on its users, exfiltrate and modify data, execute denial of services attacks, etc.

Significant efforts are being undertaken by software and data companies and researchers to protect systems, locate infections, and reverse damage inflicted by malware. Our contribution to malware analysis include the following fields:

## 7.2.1. Malware Detection

**Participants:** Axel Legay, Fabrizio Biondi, Olivier Decourbe, Mike Enescu, Thomas Given-Wilson, Annelie Heuser, Jean-Louis Lanet, Jean Quilbeuf, Alexander Zhdanov, Olivier Zendra.

Given a file or data stream, the malware detection problem consists of understanding if the file or data stream contain traces of malicious behavior. For binary executable files in particular, this requires extracting a signature of the file, so it can be compared against signatures of known clean and malicious files to determine whether the file is malicious. Binary file signatures can be divided in *syntactic* and *semantic*.

Syntactic signatures are based on properties of the file itself, like its length, hash, number and entropy of the executable and data sections, and so on. While syntactic signatures are computationally cheap to extract from binaries, it is also easy for malware creators to deploy *obfuscation* techniques that change the file's syntactic properties, hence widely mutating the signature and preventing its use for malware detection.

Semantic signatures instead are based on the binary's behavior and interactions with the system, hence are more effective at characterizing malicious files. However, they are more expensive to extract, requiring behavioral analysis and reverse-engineering of the binary. Since behavior is much harder to change than syntactic properties, against these signatures obfuscation is used to harden the file against reverse-engineering and preventing the analysis of the behavior, instead of changing it directly.

In both cases, *malware deofbuscation* is necessary to extract signatures containing actuable information that can be used to characterize the binaries as clean or malicious. Once the signatures are available, *malware classification* techniques, usually based on machine learning, are used to automatically determine whether binaries are clean or malicious starting from their signatures. Our contributions on these fields are described in the next sections.

#### 7.2.2. Malware Deobfuscation

**Participants:** Axel Legay, Fabrizio Biondi, Olivier Decourbe, Mike Enescu, Thomas Given-Wilson, Annelie Heuser, Nisrine Jafri, Jean-Louis Lanet, Jean Quilbeuf.

Given a file (usually a portable executable binary or a document supporting script macros), deobfuscation refers to the preparation of the file for the purposes of further analysis. Obfuscation techniques are specifically developed by malware creators to hinder detection reverse engineering of malicious behavior. Some of these techniques include:

- **Packing** Packing refers to the transformation of the malware code in a compressed version to be dynamically decompressed into memory and executed from there at runtime. Packing techniques are particularly effective against static analysis, since it is very difficult to determine statically the content of the unpacked memory to be executed, particularly if packing is used multiple times. The compressed code can also be encrypted, with the key being generated in a different part of the code and used by the unpacking procedure, or even transmitted remotely from a command and control (C&C) server.
- **Control Flow Flattening** This technique aims to hinder the reconstruction of the control flow of the malware. The malware's operation are divided into basic blocks, and a dispatcher function is created that calls the blocks in the correct order to execute the malicious behavior. Each block after its execution returns control to the dispatcher, so the control flow is flattened to two levels: the dispatcher above and all the basic blocks below.

To prevent reverse engineering of the dispatcher, it is often implemented with a cryptographic hash function. A more advanced variant of this techniques embed a full virtual machine with a randomly generated instruction set, a virtual program counted, and a virtual stack in the code, and uses the machine's interpreter as the dispatcher.

Virtualization is a very effective technique to prevent reverse engineering. To contrast it, we are implementing state-of-the-art devirtualization algorithms in angr, allowing it to detect and ignore the virtual machine code and retrieving the obfuscated program logic. Again, we plan to contribute our improvements to the main angr branch, thus helping the whole security community fighting virtualized malware.

**Opaque Constants and Conditionals** Reversing packing and control flow flattening techniques requires understanding of the constants and conditionals in the program, hence many techniques are deployed to obfuscate them and make them unreadable by reverse engineering techniques. Such techniques are used e.g. to obfuscate the decryption keys of packed encrypted code and the conditionals in the control flow. We have proven the efficiency of dynamic synthesis in retrieving opaque constant and conditionals, compared to the state-of-the-art approach of using SMT (Satisfiability Modulo Theories) solvers, when the input space of the opaque function is small enough. We are developing techniques based on fragmenting and analyzing by brute force the input space of opaque conditionals, and SMT constraints in general, to be integrated in SMT solvers to improve their effectiveness.

#### 7.2.3. Malware Classification

**Participants:** Axel Legay, Fabrizio Biondi, Olivier Decourbe, Mike Enescu, Thomas Given-Wilson, Annelie Heuser, Nisrine Jafri, Jean-Louis Lanet, Jean Quilbeuf.

Once malicious behavior has been located, it is essential to be able to classify the malware in its specific family to know how to disinfect the system and reverse the damage inflicted on it.

While it is rare to find an actually previously unknown malware, morphic techniques are employed by malware creators to ensure that different generations of the same malware behave differently enough than it is hard to recognize them as belonging to the same family. In particular, techniques based on the syntax of the program fails against morphic malware, since syntax can be easily changed.

To this end, semantic signatures are used to classify malware in the appropriate family. Semantic signatures capture the malware's behavior, and are thus resistant to morphic and differentiation techniques that modify the malware's syntactic signatures. We are investigating semantic signatures based on the program's System Call Dependency Graph (SCDG), which have been proven to be effective and compact enough to be used in practice. SCDGs are often extracted using a technique based on pushdown automata that is ineffective against obfuscated code; instead, we are applying concolic analysis via the angr engine to improve speed and coverage of the extraction.

Once a semantic signature has been extracted, it has to be compared against large database of known signatures representing the various malware families to classify it. The most efficient way to obtain this is to use a supervised machine learning classifier. In this approach, the classifier is trained with a large sample of signatures malware annotated with the appropriate information about the malware families, so that it can learn to quickly and automatically classify signatures in the appropriate family. Our work on machine learning classification focuses on using SCDGs as signatures. Since SCDGs are graphs, we are investigating and adapting algorithms for the machine learning classification of graphs, usually based on measures of shared subgraphs between different graphs. One of our analysis techniques relies on common subgraph extraction, with the idea that a malicious behavior characteristic of a malware family will yield a set of common subgraphs. Another approach relies on the Weisfeiler-Lehman graph kernel which uses the presence of nodes and their neighborhoods pattern to evaluate similarity between graphs. The presence or not of a given pattern becomes a feature in a subsequent machine learning analysis through random forest or SVM.

In malware detection and classification, it is fundamental to have a false positive rate (i.e. rate of cleanware classified as malware) approaching zero, otherwise the classification system will classify hundred or thousands of cleanware files as malware, making it useless in practice. To decrease the false positive rate, the classifier is also trained with a large and representative database of cleanware, so that it can discriminate between signatures of cleanware and malware with a minimal false positive rate. We use a large database of malware and cleanware to train our classifier, thus guaranteeing a high detection rate with a small false positive rate.

We have put in place a platform for malware analysis, using dedicated hardware provided by Cisco. This platform is now fully operational and receives a daily feed of suspicious binaries for analysis. Furthermore, we developed tools for maintaining our datasets of cleanware and malware binaries, run existing syntactic analysis on them. Our toolchain is able to extract SCDGs from malwares and cleanwares and apply our classification techniques on the SCDGs.

#### 7.2.4. Botnet Trojan Detection

**Participants:** Axel Legay, Fabrizio Biondi, Vesselin Bontchev, Thomas Given-Wilson, Jean Quilbeuf, Olivier Decourbe, Najah Ben Said.

Botnet trojans are a class of malware that opens a backdoor in a system and waits from further instructions from a C&C server, and possibly replicates itself somehow. A large group of systems infected by such malware is known as a botnet, and can be used by the botnet's controller to distribute spam emails (possibly carrying other malware) and perform distributed denial-of-service (DDoS) attacks. In a DDoS attack, all the systems in the botnet flood a single target with requests amounting to gigabytes or even terabytes of traffic. The target is not able to handle such traffic or to discriminate malicious request from legitimate ones, failing to provide its service.

Detecting and correctly classify botnet trojans in transit is a necessary step to be able to stop their infection. We applied our semantic classification approach on a particular family of malware, the Mirai botnet. With these experiments, we were able to confirm that the classification based on SCDG extraction and common subgraphs mining has a very low false positive rate and a high detection rate. Furthermore, our approach proved to be more accurate than detection based on syntactic signatures, without increasing the number of false positives.

### 7.2.5. Modular Automated Syntactic Signature Extraction (MASSE)

**Participants:** Axel Legay, Fabrizio Biondi, Olivier Zendra, Alexander Zhdanov, Bruno Lebon, François Déchelle.

Malware detection techniques based on syntactic signatures (or "rules") are commonly used in antivirus since their low computational cost allows them to be used on scan the files handled by the system without excessively slowing down the system. Semantic analysis techniques are relatively expensive to use, and would slow down a system significantly if used for on-access malware detection. Hence, it is common in antivirus company to use advanced semantic techniques like the SCDG-based ones we develop to detect and analyze known and unknown malware samples, and then to manually write a syntactic rule for the detection of such samples that is uploaded to the client machines.

The MASSE projects aims at providing an open-source, self-contained architecture to deploy this on a given system, company, or infrastructure, without needing to give access to the structure's data to third parties. The architecture is composed of a server executing the computationally-expensive semantic analysis, and of a number of lightweight clients performing inexpensive syntactic analysis on the client's systems. The MASSE server automatically analyzes unknown or suspicious files passing on the clients, detects the malicious ones, synthesizes syntactic signatures for them, and updates the signature databases of the clients, keeping them protected.

The MASSE server exploits modular malware analysis, supporting malware analysis modules using dynamic, static, or hybrid analysis; extracting syntactic, semantic, or hybrid signatures; using signature-based or anomaly-based detection; and any other technique the user desires, thanks to its open source malware analysis APIs. MASSE also implements pseudonymization of the signature databases, preventing an attacker to learn precisely the syntactic signatures in case some of the clients are compromised.

## 7.2.6. Malware IDS

Participants: Jean-Louis Lanet, Aurélien Palisse, Colas Le Guernic.

#### 7.2.6.1. An efficient IDS for malware detection

Ransomware is a type of malware that prevents legitimate users from accessing their machine or files and demands a payment for restoring the functionalities of the infected computer. There are two classes of ransomware: the *simple lockers*, which block the usage of the computer, and *cryptors*, that encrypt files on the computer. In the case of encryption-based ransomware, the user data can only be restored with the secret key(s) used during the attack if the key is provided by the attacker.

Detecting a malware can use two options:

• The system knows the features of the malware. Features can be structural information: n-gram or graph isomorphism, or behavioral information: APIs call or system calls. Exact pattern matching algorithm or approximative algorithm (Machine learning) can be used. This approach is known as signature based and can only detect known patterns.

• The system knows its correct behavior. Any deviation of this model leads to the detection of hostile programs. This approach can detect any new attack, it does not rely on a model of the bad behavior but on the model of the correct behavior. This approach is also known as IDS (Intrusion Detection System).

In [45], [34] we apply this technique to detect malware at run time (EPS: End Point Solution). Our first solution is based on the dynamic analysis of the data transformation by the program. We propose to monitor file activity. Since it has already been proven a valid approach in terms of detection, our main goal in is to show that a good detection rate can be achieved with little to no impact on system performances. To this end, we limit our monitoring to a minimum. In order to reduce the impact on detection with a low rate of false positive, we use the chi-square goodness-of-fit test instead of Shannon entropy (*i.e.*, sensitive to compressed chunks of data). We also achieve system completeness and fine granularity by monitoring the whole file system for all userland threads. In order to evaluate our prototype implementation, Data Aware Defense (DaD), under realistic conditions, we used the bare-metal analysis platform of the LHS, Malware - O - Matic (MoM), and ran it on a large and heterogeneous (compared to the literature) live ransomware collection. We used *de facto* industry standard benchmarks to get a pertinent and reproducible assessment of the performance penalties. A second model of the correct behavior with better results has been developed (patent pending).

Our countermeasure is efficient and can be deployed on Windows 7/10 machines with a reasonable performance hit, with an average delay of 12  $\mu$ s per write operation on disk, a few hundred times smaller than previous approaches. Our extensive experiments show that the more sophisticated ransomware already use mimicry attacks. However we successfully detect 99.37 % of the samples with at most 70 MB lost per sample's threads in 90% of cases and less than 7 MB in 70% of cases. Its speed and low negative rate makes it a good candidate as a first line of defense. Once a thread is deemed malicious, instead of blocking disk accesses, other more costly metrics can be used to improve the false positive rate without impacting performance, since it would not be computed for all other threads.

#### 7.2.7. Papers

This section gathers papers that are results common to all sections above pertaining to Axis 2.

- [51] (C) The largest DDoS attacks in history have been executed by devices controlled by the Mirai botnet trojan. To prevent Mirai from spreading, this paper presents and evaluates techniques to classify binary samples as Mirai based on their syntactic and semantic properties. Syntactic malware detection is shown to have a good detection rate and no false positives, but to be very easy to circumvent. Semantic malware detection is resistant to simple obfuscation and has better detection rate than syntactic detection, while keeping false positives to zero. This paper demonstrates these results, and concludes by showing how to combine syntactic and semantic analysis techniques for the detection of Mirai.
- [19] (C) We present the MASSE architecture, a YARA-based open source client-server malware detection platform. MASSE includes highly effective automated syntactic malware detection rule generation for the clients based on a server-side modular malware detection system. Multiple techniques are used to make MASSE effective at detecting malware while keeping it from disrupting users and hindering reverse-engineering of its malware analysis by malware creators.
- [4] (J) Control flow obfuscation techniques can be used to hinder software reverse-engineering. Symbolic analysis can counteract these techniques, but only if they can analyze obfuscated conditional statements. We evaluate the use of dynamic synthesis to complement symbolic analysis in the analysis of obfuscated conditionals. We test this approach on the taint-analysis-resistant Mixed Boolean Arithmetics (MBA) obfuscation method that is commonly used to obfuscate and randomly diversify statements. We experimentally ascertain the practical feasibility of MBA obfuscation. We study using SMT-based approaches with different state-of-the-art SMT solvers to counteract MBA obfuscation, and we show how targeted algebraic simplification can greatly reduce the analysis time. We show that synthesis-based deobfuscation is more effective than current SMT-based deobfuscation algorithms, thus proposing a synthesis-based attacker model to complement existing attacker models.

## 7.3. Results for Axis 3: Building a secure network stack

#### 7.3.1. Privacy-Preserving Abuse Detection in Future Decentralised Online Social Networks

Participants: Jeffrey Burdges, Alvaro Garcia Recuero, Christian Grothoff.

Future online social networks need to not only protect sensitive data of their users, but also protect them from abusive behavior coming from malicious participants in the network. We investigated the use of supervised learning techniques to detect abusive behavior and describe privacy-preserving protocols to compute the feature set required by abuse classification algorithms in a secure and privacy-preserving way. While our method is not yet fully resilient against a strong adaptive adversary, our evaluation suggests that it will be useful to detect abusive behavior with a minimal impact on privacy.

Our results show how to combine local knowledge with private set intersection and union cardinality protocols (with masking of BLS signature to protect identity of signers/subscribers) to privately derive feature values from users in OSNs. Given an adaptive adversary that would be able to manipulate most features we propose in our supervised learning approach, it is surprising that with just three features resistant to adversarial manipulation, the algorithms still provide useful classifications.

This work was originally presented at DPM 2016 [63] and expanded upon in Álvaro García-Recuero's PhD thesis [1].

## 7.3.2. Fog of Trust

Participants: Jeffrey Burdges, Christian Grothoff.

The Web of Trust (WoT) used traditionally used by tools for private communication such as PGP is used to to validate individual links between participants. Using the WoT, however, leaks meta data, such that users must opt-in for it – exposing themselves to risks of privacy loss. We proposed a new method, the Fog of Trust (FoT), which uses the privacy-preserving set intersection cardinality protocol originally used in our work on abuse detection in online social networks, to support this critical step of public key verification via collaboration. In the FoT, the social relationships — which are used to verify public keys – remain hidden. This allows keys to be verified via trusted intermediaries that were established beforehand, without the need to verify each individual new contact using Trustwords. Consequently, FoT will can the same functionality as the WoT without its drawbacks to privacy.

## 7.3.3. Cell tower privacy

Participants: Christian Grothoff, Neal Walfield.

Context-aware applications are programs that are able to improve their performance by adapting to the current conditions, which include the user's behavior, networking conditions, and charging opportunities. In many cases, the user's location is an excellent predictor of the context. Thus, by predicting the user's future location, we can predict the future conditions. In this work, we developed techniques to identify and predict the user's location over the next 24 hours with a minimum median accuracy of 82 results include our observation that cell phones sample the towers in their vicinity, which makes cell towers as-is inappropriate for use as landmarks. Motivated by this observation, we developed two techniques for processing the cell tower traces so that landmarks more closely correspond to locations, and cell tower transitions more closely correspond to user movement. We developed a prediction engine, which is based on simple sampling distributions of the form f(t,c), where t is the predicted tower, and c is a set of conditions. The conditions that we considered include the time of the day, the day of the week, the current regime, and the current tower. Our family of algorithms, called TomorrowToday, achieves 89% prediction precision across all prediction trials for predictions 30 minutes in the future. This decreases slowly for predictions further in the future, and levels off for predictions approximately 4 hours in the future, at which point we achieve 82% prediction precision across all prediction trials up to 24 hours in the future. This represents a significant improvement over NextPlace, a well-cited prediction algorithm based on non-linear time series, which achieves appropriately 80% prediction precision (self reported) for predictions 30 minutes in the future, but, unlike our predictors, which try all prediction attempts, NextPlace only attempts 7% of the prediction trials on our data set [67].

#### 7.3.4. Taler protocol improvements

Participants: Jeffrey Burdges, Florian Dold, Christian Grothoff, Marcello Stanisci.

We started modeling the Taler protocol in the framework of Provable Security, precisely defining the formal meaning of income transparency, fairness, anonymity and unforgeablity as security games. The resulting definitions and security proofs allow a more precise statement of the security of Taler in relation to the security assumptions that are being made.

The implementation of the wallet module now supports the full Taler protocol, including the refresh operation for highly efficient and privacy-preserving change.

In addition to improving the stability of the implementation of all Taler components, we added new features to the protocol that (1) allow refunds from merchants without violating privacy and (2) allow merchants to do "customer tipping", which transfers money from merchants directly to customers' wallets as a reward for doing actions on their website.

### 7.3.5. Mix Networking

Participants: Jeffrey Burdges, Christian Grothoff.

We have begun implementing our ratcheting scheme for providing hybrid post-quantum and forward security to the Sphinx mix network packet format. We also began collaborating with the Panoramix project and LEAP to help resolve numerous practical challenges to deploying a mix network. We shall speak about this ongoing work at the Chaos Computer Club's annual congress 34c3 in December 2017.

## 7.4. Other research results

### 7.4.1. Privacy and Security: Information-Theoretical Quantification of Security Properties

**Participants:** Axel Legay, Fabrizio Biondi, Olivier Zendra, Thomas Given-Wilson, Annelie Heuser, Sean Sedwards, Jean Quilbeuf, Mike Enescu.

Information theory provides a powerful quantitative approach to measuring security and privacy properties of systems. By measuring the *information leakage* of a system security properties can be quantified, validated, or falsified. When security concerns are non-binary, information theoretic measures can quantify exactly how much information is leaked. The knowledge of such information is strategic in the developments of component-based systems.

The quantitative information-theoretical approach to security models the correlation between the secret information of the system and the output that the system produces. Such output can be observed by the attacker, and the attacker tries to infer the value of the secret information by combining this information with their prior knowledge of the system.

Armed with the produced output of the system, the attacker tries to infer information about the secret information that produced the output. The quantitative analysis we consider defines and computes how much information the attacker can expect to infer (typically measured in bits). This expected leakage of bits is the information leakage of the system. This is computed by symbolically exploring the code to be analyzed, and using the symbolic constraints accumulated over the output together with a model counting algorithm to quantify the leakage.

The quantitative approach generalizes the qualitative approach and thus provides superior analysis. In particular, a system respects non-interference if and only if its leakage is equal to zero. In practice very few systems respect non-interference, and for those that don't it is imperative to be able to distinguish between the systems leaking very small amounts of secret information and systems leaking a significant amount of secret information, since only the latter are considered to pose a security vulnerability to the system. While quantitative leakage computation is a powerful technique to detect security vulnerabilities, computing the leakage of complex programs written in low-level languages is a hard and computationally intensive task. The most common language for low-level implementation of security protocols is C, due to its efficiency, hence much of the effort in developing tools to detect vulnerabilities in source code focus on C. Recently, we have improved the state of the art in leakage quantification from C programs by proposing the usage of approximated model counting instead of precise model counting. We have shown how the approximation can improve the efficiency of leakage quantification by orders of magnitude against a logarithmic decrease in the precision of the result, often producing the same result as precise model counters much faster, and often being able to analyze cases where precise model counters would have failed. We demonstrated this technique by providing the first quantitative leakage analysis of the C code of the Heartbleed bug, showing that our technique can detect the bug in the code.

A different but equally interesting approach is followed by our new HyLEak tool. HyLeak is also able to analyze a system and compute its information leakage, i.e. the amount of information that an observer would gain by about the value of system's secret by observing its output. Contrarily to other techniques, HyLeak can analyze randomized systems, and correctly distinguish between the randomness injected in the system and the uncertainty on the secret value. This allows HyLeak to be used both on systems with explicit randomization and systems that depend on stochastic properties, like cyber-physical systems.

HyLeak uses static code analysis to divide the system to be analyzed in components. For each component, HyLeak evaluates whether it is more convenient to analyze the component using precise or statistical analysis. Each component is analyzed with the most appropriate strategy, and then the results for all components are combined together and information leakage is estimated.

The hybrid approach provides better results than both the precise and the statistical ones in terms of computation time and precision of the result. Also, it bridges the gap between cheap but imprecise statistical techniques and precise but expensive formal techniques, allowing the user to control the required precision of the result according to the computation time they have available. We evaluated HyLeak against QUAIL's precise approach and the statiatical approach implemented in the LeakWatch tool, showing that HyLeak outperforms them both. HyLeak is open source and available at https://project.inria.fr/hyleak/

Applied to shared-key cryptosystems, the information-theoretical approach allows precise reasoning about the information leakage of any secret information in the system including, the key, and the message. Recent work on max-equivocation has generalised perfect secrecy and shown the maximum achievable theoretic bounds for the security of the key and message. Achieving these theoretic maximal bounds has been proven to be achievable by Apollonian Cell Encoders (ACEs). ACEs not only allow the maximum security possible in a shared-key cryptosystem, but also allow for infinite key reuse when the key has less entropy than the message. Further, ACEs are straightforward to construct and have a compact representation making them feasible to use in practice.

Another application is to use information leakage to reason about leakage through shared resources, representing various side-channel attacks. Developmens here allow for the formalising of the leakage model through shared resources, and quantifying how significant the leakage can be. This improves on the state-of-the-art that uses only qualified leakage, and so can be precise about how much is leakage through a shared resource. Such quantification of leakage allows for scheduling of the shared resource to exploit this information to minimise leakage. Such minimisation of leakage allows for scheduling and utilisation of resources that would fail a simple quanlified test, providing solutions when prior state-of-the-art would claim impossibility. Further, a reasoned trade-off can be made between acceptable leakage and utility of the shared resource, allowing solutions that are acceptable even if not perfect.

[53] (C; submitted) Preserving privacy of private communication against an attacker is a fundamental concern of computer science security. Unconditional encryption considers the case where an attacker has unlimited computational power, hence no complexity result can be relied upon for encryption. Optimality criteria are defined for the best possible encryption over a general collection of entropy measures. This paper introduces Apollonian cell encoders, a class of shared-key cryptosystems that are proven to be universally optimal. In addition to the highest possible security for the message,

Apollonian cell encoders prove to have perfect secrecy on their key allowing unlimited key reuse. Conditions for the existence of Apollonian cell encoders are presented, as well as a constructive proof. Further, a compact representation of Apollonian cell encoders is presented, allowing for practical implementation.

- [18] (C) High-security processes have to load confidential information into shared resources as part of their operation. This confidential information may be leaked (directly or indirectly) to low-security processes via the shared resource. This paper considers leakage from high-security to low-security processes from the perspective of scheduling. The workflow model is here extended to support preemption, security levels, and leakage. Formalization of leakage properties is then built upon this extended model, allowing formal reasoning about the security of schedulers. Several heuristics are presented in the form of compositional preprocessors and postprocessors as part of a more general scheduling approach. The effectiveness of such heuristics are evaluated experimentally, showing them to achieve significantly better schedulability than the state of the art. Modeling of leakage from cache attacks is presented as a case study.
- [52] (C) Quantitative information flow measurement techniques have been proven to be successful in detecting leakage of confidential information from programs. Modern approaches are based on formal methods, relying on program analysis to produce a SAT formula representing the program's behavior, and model counting to measure the possible information flow. However, while program analysis scales to large codebases like the OpenSSL project, the formulas produced are too complex for analysis with precise model counting. In this paper we use the approximate model counter ApproxMC2 to quantify information flow. We show that ApproxMC2 is able to provide a large performance increase for a very small loss of precision, allowing the analysis of SAT formulas produced from complex code. We call the resulting technique ApproxFlow and test it on a large set of benchmarks against the state of the art. Finally, we show that ApproxFlow can evaluate the leakage incurred by the Heartbleed OpenSSL bug, contrarily to the state of the art.
- [20] (C) We present HyLeak, a tool for reasoning about the quantity of information leakage in programs. The tool takes as input the source code of a program and analyzes it to estimate the amount of leaked information measured by mutual information. The leakage estimation is mainly based on a hybrid method that combines precise program analysis with statistical analysis using stochastic program simulation. This way, the tool combines the best of both symbolic and randomized techniques to provide more accurate estimates with cheaper analysis, in comparison with the previous tools using one of the analysis methods alone. HyLeak is publicly available and is able to evaluate the information leakage of randomized programs, even when the secret domain is large. We demonstrate with examples that HyLeaks has the best performance among the tools that are able to analyze randomized programs with similarly high precision of estimates.
- [54] (J; submitted) Analysis of a probabilistic system often requires to learn the joint probability distribution of its random variables. The computation of the exact distribution is usually an exhaustive precise analysis on all executions of the system. To avoid the high computational cost of such an exhaustive search, statistical analysis has been studied to efficiently obtain approximate estimates by analyzing only a small but representative subset of the system's behavior. In this paper we propose a hybrid statistical estimation method that combines precise and statistical analyses to estimate mutual information, Shannon entropy, and conditional entropy, together with their confidence intervals. We show how to combine the analyses on different components of the system with different accuracy to obtain an estimate for the whole system. The new method performs weighted statistical analysis with different sample sizes over different components and dynamically finds their optimal sample sizes. Moreover it can reduce sample sizes by using prior knowledge about systems and a new abstraction-then-sampling technique based on qualitative analysis. To apply the method to the source code of a system, we show how to decompose the code into components and to determine the analysis method for each component by overviewing the implementation of those techniques in HyLeak tool. We demonstrate with case studies that the new method outperforms the state of the art in quantifying information leakage.

#### 7.4.2. Security for therapeutical environments

Participants: Axel Legay, Olivier Zendra, Thomas Given-Wilson, Sean Sedwards.

This work is done in the context of the ACANTO EU project. We aim at helping develop robotic assistants to aid mobility of mobility-impaired and elderly adults. These robotic assistants provide a variety of support to their users, including: navigational assistance, social networking, social activity planning, therapeutic regime support, and diagnostic support. In Tamis, we focus on navigational assistance and social activities, as together they yield an interesting challenge in human robot interaction. The goal is to help groups of users navigate in a potentially busy dynamic environment, while also maintaining social group cohesion.

A robotic assistant has been developed before in the DALi project, acting selfishly to ensure the safe navigation of a single user. This was achieved by using the social force model and statistical model checking in a reactive planner that frequently replanned and made immediate navigational suggestions to the user. The key operational loop of this solution was to: observe the environment, model the agents in the environment in the social force model, give safety constraints for the user, and then use statistical model checking to find the optimal next move for the user.

Generalising to groups of users poses several significant difficulties. Computationally, the challenge is exponential in the number of users, considering all their possible navigational choices. Incomplete information is normal, since sensors are distributed between robotic assistants and the environment, and communication may fail, leading to different robots having different knowledge of the environment. Maintaining group cohesion is non-trivial, since group composition and position are dynamic and, unlike swarm robotics, no group member can be abandoned. Frequent replanning is necessary since there is minimal control over the users' actions, which may include ignoring the advise of the robotic assistant

The solution we designed is to abstract away from individual users in favour of groups. This refines the prior solution for a single user. Sensor information is used to obtain traces that provide behavioural information about users and pedestrians in the environment. These traces are clustered into groups that capture both location and motion behaviour. The groups are used as the social particles in the social force model, with parameters adjusted to account for group dynamics. Statistical model checking is used to find the optimal next move for the group containing the user, and the navigation for the optimal next move is displayed to the user. The effectiveness of the group abstraction mechanisms use in this refined algorithm are validated on the BIWI walking pedestrians dataset. This shows they operate correctly and effectively, even improving over human annotations, on real world data of pedestrians in a chaotic environment.

- [27] (C) People with impaired physical and mental ability often find it challenging to negotiate crowded or unfamiliar environments, leading to a vicious cycle of deteriorating mobility and sociability. To address this issue the ACANTO project is developing a robotic assistant that allows its users to engage in therapeutic group social activities, building on work done in the DALi project. Key components of the ACANTO technology are social networking and group motion planning, both of which entail the sharing and broadcasting of information. Given that the system may also make use of medical records, it is clear that the issues of security, privacy, and trust are of supreme importance to ACANTO.
- [58] (C; submitted) The ACANTO project is developing robotic assistants to aid the mobility and recovery of mobility-impaired and older adults. One key feature of the project's robotic assistants is aiding with navigation in chaotic environments. Prior work has solved this for a single user with a single robot, however for therapeutic outcomes ACANTO supports social groups and group activities. Thus these robotic assistants must be able to efficiently support groups of users walking together. This requires an efficient navigation solution that can handle large numbers of users, maintain (de-facto) group cohesion despite unpredictable behaviours, and operate rapidly on embedded devices. We address these challenges by: using sensor information to develop behavioural traces, clustering traces to determine groups, modeling the groups using the social force model, and finding an optimal navigation solution using statistical model checking. The new components of this solution are validated on the ETH Zürich dataset of pedestrians in an open environment.

## 7.4.3. Mobile air pollution sensor platform for smart-cities

Participant: Laurent Morin.

This work is organized and coordinated by the Chaire "mobilité dans une ville durable" and financed by the Foundation of Rennes 1 (https://fondation.univ-rennes1.fr/)

The purpose of this work is to design and experiment a mobile pollution sensor platform for Smart-Cities in Rennes.

The platform is integrated in the project ROAD (Rennes Open Access to Data) proposing to development of mobile systems operating the collection and the management of open data in Rennes for a future development of a smart-city. The collaboration is part of an ecosystem developed by the Chair "mobilité dans une ville durable" via the production of multiple experimentations in the city.

In the ROAD project context, the air quality in the city has been identified as one of the major challenge. Air quality improvement can only be achieved with a citizen and political full cooperation and involvement. This experimentation aims at providing an end-to-end urban platform that extends current practices in air quality measurements and allows citizens and policy makers to obtain the data and make informed decisions.

The mobile air pollution sensor platform for smart-cities proposes a innovative IoT architecture introducing the deployment of a small set of advanced and cost-effective sensors around a balanced high-performance/low-power compute unit inside a mobile agent in the city. The compute unit will have to provide the necessary computation power needed to produce advanced analysis and the security management on-site (integrity, authentication, ...).

The mobile sensor platform developments partially started in July 2017, and accelerated in October for a real deployment in buses in 2018. During this period, the core system of the platform was designed, adapted, and partially implemented to offer an operational prototype. This year lead to the design of a suitcase containing a self-sufficient measurement system: a main compute unit, its power supply and power management, and a set of satellite pollution sensors. This achievement was disseminated to the Rennes ecosystem (Rennes Atalante, Rennes Métropole, Inria) through the participation to several meetings and exhibitions.

# 8. Bilateral Contracts and Grants with Industry

# 8.1. Bilateral Contracts with Industry

• CISCO (http://www.cisco.com) contract (2017–2022) to work on graph analysis of malware

# 8.2. Bilateral Grants with Industry

- CISCO (http://www.cisco.com) one grant (2016–2019) to work on semantical analysis of malware
- Thales (https://www.thalesgroup.com) one CIFRE (2016–2019) to work on verification of communication protocols, one grant (2018–2019) to work on learning algorithms
- Oberthur Technologies (http://www.oberthur.com/) one grant (2016–2020) to work on fuzzing and fault injection
- Secure IC (http://www.secure-ic.com/), one CIFRE (2017–2020) to work on post-quantum cryptography

# 9. Partnerships and Cooperations

# 9.1. Regional Initiatives

• ARED grant for Lamine Nouredine and Florian Dolt

- Postdocs grants for Najah Ben Said, Jeffrey Paul Burdges, Ronan Lashermes, Ludovic Claudepierre
- Starting Grant for hardware for Annelie Heuser from Rennes Metropole
- Software developer grant for Laurent Morin from "Chaire Mobilité dans une ville durable" (mobility in a sustainable city) by Fondation Université Rennes 1

# 9.2. National Initiatives

## 9.2.1. ANR

- ANR MALTHY, Méthodes ALgébriques pour la vérification de modèles Temporisés et HYbrides, Thao Dang, 4 years, Inria and VISEO and CEA and VERIMAG
- ANR COGITO, Runtime Code Generation to Secure Devices, 3 years, Inria and CEA and ENSMSE and XLIM.

## 9.2.2. DGA

- PhD grant for Nisrine Jafri (2016–2019),
- PhD grant for Aurélien Palisse (2016–2019),
- PhD grant for Alexandre Gonzalves (2016–2019),
- PhD grant for Olivier Decourbe (2017–2020),
- PhD grant for Alexandre Zdhanov (2017–2020)

#### 9.2.3. Autres

• INS2I JCJC grant for Axel Legay, Annelie Heuser, Fabrizio Biondi.

## 9.3. European Initiatives

## 9.3.1. FP7 & H2020 Projects

9.3.1.1. ACANTO

Title: ACANTO: A CyberphusicAl social NeTwOrk using robot friends

Program: H2020

Duration: February 2015 - July 2018

Coordinator: Universita di Trento

Partners:

Atos Spain (Spain), Envitel Tecnologia Y Control S.A. (Spain), Foundation for Research and Technology Hellas (Greece), Servicio Madrileno Delud (Spain), Siemens Aktiengesellschaft Oesterreich (Austria), Telecom Italia S.P.A (Italy), Universita' Degli Studi di Siena (Italy), Universita Degli Studi di Trento (Italy), University of Northumbria At Newcastle. (United Kingdom)

#### Inria contact: Axel Legay

Despite its recognised benefits, most older adults do not engage in a regular physical activity. The ACANTO project proposes a friendly robot walker (the FriWalk) that will abate a some of the most important barriers to this healthy behaviour. The FriWalk revisits the notion of robotic walking assistants and evolves it towards an activity vehicle. The execution of a programme of physical training is embedded within familiar and compelling every-day activities. The FriWalk operates as a personal trainer triggering the user actions and monitoring their impact on the physical and mental well-being. It offers cognitive and emotional support for navigation pinpointing risk situations in the environment and understanding the social context. It supports coordinated motion with other FriWalks for group activities. The FriWalk combines low cost and advanced features, thanks to its reliance on a cloud of services that increase its computing power and interconnect it to other assisted

living devices. Very innovative is its ability to collect observations on the user preferred behaviours, which are consolidated in a user profile and used for recommendation of future activities. In this way, the FriWalk operates as a gateway toward a CyberPhysical Social Network (CPSN), which is an important contribution of the project. The CPSN is at the basis of a recommendation system in which users' profiles are created, combined into 'circles' and matched with the opportunity offered by the environment to generate recommendations for activities to be executed with the FriWalk support. The permanent connection between users and CPSN is secured by the FriPad, a tablet with a specifically designed user interface. The CPSN creates a community of users, relatives and therapists, who can enter prescriptions on the user and receive information on her/his state. Users are involved in a large number in all the phases of the system development and an extensive validation is carried out at the end.

Axel Legay and Olivier Zendra are the permanent researchers of Tamis involved in this project. The project supports two postdocs in Tamis.

#### 9.3.1.2. DIVIDEND

Title: DIVIDEND: Distributed Heterogeneous Vertically IntegrateD Energy Efficient Data centres Program: CHIST-ERA 2013

Duration: 10/2014 - 10/2016 (extended 10/2017)

Coordinator: University of Edinburgh (UK)

Partners:

École Normale Supérieure de Paris, Département d'Informatique (France); Inria (France); Ecole Polytechnique Fédérale de Lausanne, Computer & Communication Sciences (Switzerland); Queen's University of Belfast, School of Electronics, Electrical Engineering and Computer Science, Belfast (UK); University of Edinburgh, Scotland, (UK); University of Lancaster, School of Computing and Communications (UK); University Politehnica Timisoara, Department of Computer Engineering (Romania)

Inria contact: Albert Cohen

The DIVIDEND project (http://www.chistera.eu/projects/dividend) attacks the data centre energy efficiency bottleneck through vertical integration, specialization, and cross-layer optimization. Our vision is to present heterogeneous data centres, combining CPUs, GPUs, and task-specific accelerators, as a unified entity to the application developer and let the runtime optimize the utilization of the system resources during task execution. DIVIDEND embraces heterogeneity to dramatically lower the energy per task through extensive hardware specialization while maintaining the ease of programmability of a homogeneous architecture. To lower communication latency and energy, DIVIDEND refers a lean point-to-point messaging fabric over complex connection-oriented network protocols. DIVIDEND addresses the programmability challenge by adapting and extending the industry-led heterogeneous systems architecture programming language and runtime initiative to account for energy awareness and data movement. DIVIDEND provides for a cross-layer energy optimization framework via a set of APIs for energy accounting and feedback between hardware, compilation, runtime, and application layers. The DIVIDEND project will usher in a new class of vertically integrated data centres and will take a first stab at resolving the energy crisis by improving the power usage effectiveness of data centres.

Contributions of Inria in the project addresses the development of an energy aware distributed heterogeneous system (distributed HSA) between data center applications and HSA accelerators. It includes the design of a common API able to interface two tasks: the monitoring of the energy consumption, and the management of distributed heterogeneous hardware at a data center scale. The project ended by a project review the 23th March 2017, and the last contributions to the project ended the 30th September 2017.

One of the main contribution is the design of and energy-aware distributed heterogeneous system architecture framework (D-HSA) built using the combination of three major levels: the hardware

platform based on an aggregation of HSA compliant devices, the system level based on device drivers and energy monitoring libraries, and finally the application layer using an extension of standard OpenCL programming model. This OpenCL extension is proposed as the main API for the energyaware distributed HSA, and was made available for the tools and applications developed in the project.

A second contribution is the specification and the implementation of a distributed extension of the standard HSA Runtime API, and its functional validation on a basic system. The extension integrates the discovery, the management, and the execution of kernel computations on remote HSA agents in a distributed environment. The validation is based on an implementation using the Message Passing Interface (MPI) standard on an HSA compliant AMD machine. The Distributed HSA extension proposed offers a fully functional API for managing remote and distributed HSA agents, but at the cost of a limitation of the capability of the D-HSA system: the standard HSA memory model, based essentially on a coherent shared memory, is not supported for distributed HSA agents. As a primary implementation, focusing on a functional support of the new D-HSA verbs, this works tend to demonstrate that the extension is light and easy-to-use for a set of examples.

Laurent Morin from Tamis is involved in this project

#### 9.3.1.3. EMC2

Title: Embedded Multi-Core Systems for Mixed Criticality Applications in Dynamic and Changeable Real-Time Environments

Program: FP7

Duration: April 2014 - March 2017

Coordinator: Infineon Technologies

Partners:

Aicas (Germany) Avl Software and Functions (Germany), Denso Automotive Deutschland (Germany), Elektrobit Automotive (Germany), Evision Systems (Germany), Nxp Semiconductors Germany (Germany), Tttech Computertechnik (Austria), "kompetenzzentrum - Das Virtuelle Fahrzeug, Forschungsgesellschaft Mbh" (Austria), Frequentis (Austria), Thales Austria (Austria), Blueice Bvba (Belgium), Freescale Polovodice Ceska Republika Sro (Czech Republic), Sysgo Sro (Czech Republic), Silkan Rt (France), "united Technologies Research Centre Ireland," (Ireland), Mbda Italia Spa (Italy), Fornebu Consulting As (Norway), Westerngeco As (Norway), Simula Research Laboratory As (Norway), Ixion Industry and Aerospace SI (Spain), Visure Solutions SI (Spain), Seven Solutions SI (Spain), Telvent Energia (Spain), Instituto Tecnologico de Informatica (Spain), Ambar Telecomunicaciones SI (Spain), Sics Swedish Ict (Sweden), Arcticus Systems (Sweden), Arccore (Sweden), Xdin Stockholm (Sweden), Systemite (Sweden), Stichting Imec Nederland (Netherlands), Tomtom International Bv (Netherlands), Infineon Technologies Uk Ltd (United Kingdom), Sundance Multiprocessor Technology Ltd (United Kingdom), Systonomy (United Kingdom), Ensilica Ltd (United Kingdom), Test and Verification Solutions Ltd (United Kingdom), Abb (Sweden), Ait Austrian Institute of Technology (Austria), Alenia Aermacchi Spa (Italy), Avl List (Austria), Airbus Defence and Space (Germany), Bayerische Motoren Werke Aktiengesellschaft (Germany), Vysoke Uceni Technicke V Brne (Czech Republic), Commissariat A L Energie Atomique et Aux Energies Alternatives (France), Consorzio Interuniversitario Nazionale Per l'Informatica (Italy), Centro Ricerche Fiat (Italy), Critical Software (Portugal), Chalmers Tekniska Hoegskola (Sweden), Danfoss Power Electronics As (Denmark), Danmarks Tekniske Universitet (Denmark), Ericsson (Sweden), Fraunhofer-Gesellschaft Zur Foerderung Der Angewandten Forschung E.V (Germany), Hi Iberia Ingenieria Y Proyectos SI (Spain), Harokopio University (Greece), Infineon Technologies Austria (Austria), Institut Mikroelektronickych Aplikaci S.R.O. (Czech Republic), Inesc Id - Instituto de Engenharia de Sistemas E

Computadores, Investigacao E Desenvolvimento Em Lisboa (Portugal), Infineon Technologies (Germany), Integrasys (Spain), Instituto Superior de Engenharia Do Porto (Portugal), Kungliga Tekniska Hoegskolan (Sweden), Lulea Tekniska Universitet (Sweden), Magillem Design Servicess (France), Nxp Semiconductors Netherlands Bv (Netherlands), Offis E.V. (Germany), Philips Medical Systems Nederland Bv (Netherlands), Politecnico di Torino (Italy), Quobis Networks Sl (Spain), Rockwell Collins France (France), Rigas Tehniska Universitate (Latvia), Selex Es Spa (Italy), Siemens Aktiengesellschaft (Germany), Systematic Paris Region Association (France), Sysgo (Germany), Thales Alenia Space Italia Spa (Italy), "thales Alenia Space Espana," (Spain), Technolution B.V. (Netherlands), Fundacion Tecnalia Research & Innovation (Spain), Thales Communications & Securitys (France), Thales Avionicss (France), Thales (France), Nederlandse Organisatie Voor Toegepast Natuurwetenschappelijk Onderzoek Tno (Netherlands), Technische Universitat Braunschweig (Germany), Technische Universiteit Delft (Netherlands), Technische Universitat Dortmund (Germany), Technische Universitaet Kaiserslautern (Germany), Technische Universitaet Wien (Austria), Technische Universiteit Eindhoven (Netherlands), Universita Degli Studi di l'aquila (Italy), Universita Degli Studi di Genova (Italy), The University of Manchester (United Kingdom), University of Bristol (United Kingdom), University of Limerick (Ireland), "ustav Teorie Informace A Automatizace Av Cr, V.V.I." (Czech Republic), Universitetet I Oslo (Norway), Vector Fabrics Bv (Netherlands), Volvo Technology (Sweden)

Inria contact: Albert Cohen and Axel Legay

Embedded systems are the key innovation driver to improve almost all mechatronic products with cheaper and even new functionalities. Furthermore, they strongly support today's information society as inter-system communication enabler. Consequently boundaries of application domains are alleviated and ad-hoc connections and interoperability play an increasing role. At the same time, multi-core and many-core computing platforms are becoming available on the market and provide a breakthrough for system (and application) integration. A major industrial challenge arises facing (cost) efficient integration of different applications with different levels of safety and security on a single computing platform in an open context. The objective of the EMC2 project (Embedded multicore systems for mixed criticality applications in dynamic and changeable real-time environments) is to foster these changes through an innovative and sustainable service-oriented architecture approach for mixed criticality applications in dynamic and changeable real-time environments. The EMC2 project focuses on the industrialization of European research outcomes and builds on the results of previous ARTEMIS, European and National projects. It provides the paradigm shift to a new and sustainable system architecture which is suitable to handle open dynamic systems. EMC2 is part of the European Embedded Systems industry strategy to maintain its leading edge position by providing solutions for: . Dynamic Adaptability in Open Systems . Utilization of expensive system features only as Service-on-Demand in order to reduce the overall system cost. . Handling of mixed criticality applications under real-time conditions . Scalability and utmost flexibility . Full scale deployment and management of integrated tool chains, through the entire lifecycle Approved by ARTEMIS-JU on 12/12/2013 for EoN. Minor mistakes and typos corrected by the Coordinator, finally approved by ARTEMIS-JU on 24/01/2014. Amendment 1 changes approved by ECSEL-JU on 31/03/2015.

The permanent members of Tamis who are involved are Axel Legay and Olivier Zendra. The project was initiated during the lifetime of the ESTASYS.Inria team.

#### 9.3.1.4. ENABLE-S3

Title: ENABLE-S3: European Initiative to Enable Validation for Highly Automated Safe and Secure Systems Program: H2020 Duration: 05/2016 - 04/2019 Coordinator: Avl List Gmbh (Austria)

#### Partners:

Aalborg Universitet (Denmark); Airbus Defence And Space Gmbh (Germany); Ait Austrian Institute Of Technology Gmbh (Austria); Avl Deutschland Gmbh (Germany); Avl Software And Functions Gmbh (Germany); Btc Embedded Systems Ag (Germany); Cavotec Germany Gmbh (Germany); Creanex Oy( Finland); Ceske Vysoke Uceni Technicke V Praze (Czech Republic); Deutsches Zentrum Fuer Luft - Und Raumfahrt Ev (Germany); Denso Automotive Deutschland Gmbh (Germany); Dr. Steffan Datentechnik Gmbh (Austria); Danmarks Tekniske Universitet (Denmark); Evidence Srl (Italy); Stiftung Fzi Forschungszentrum Informatik Am Karlsruher Institut Fur Technologie (Germany); Gmv Aerospace And Defence Sa (Spain); Gmvis Skysoft Sa (Portugal); Politechnika Gdanska (Poland); Hella Aglaia Mobile Vision Gmbh (Germany); Ibm Ireland Limited (Ireland); Interuniversitair Micro-Electronica Centrum (Belgium); Iminds (Belgium); Institut National De Recherche Eninformatique Et Automatique (France); Instituto Superior De Engenharia Do Porto (Portugal); Instituto Tecnologico De Informatica (Spain); Ixion Industry And Aerospace SI (Spain); Universitat Linz (Austria); Linz Center Of Mechatronics Gmbh (Austria); Magillem Design Services Sas (France); Magneti Marelli S.P.A. (Italy); Microeletronica Maser Slspain); Mdal (France); Model Engineering Solutions Gmbhgermany); Magna Steyr Engineering Ag & Co Kg (Austria); Nabto Aps (Denmark); Navtor As (Norway); Nm Robotic Gmbh (Austria); Nxp Semiconductors Germany Gmbh(Germany); Offis E.V.(Germany); Philips Medical Systems Nederland Bvnetherlands); Rohde & Schwarz Gmbh&Co Kommanditgesellschaft(Germany); Reden B.V. (Netherlands); Renault Sas (France); Rugged Tooling Oyfinland); Serva Transport Systems Gmbh(Germany); Siemens Industry Software Nvbelgium); University Of Southampton (Uk); Safetrans E.V. (Germany); Thales Alenia Space Espana, Saspain); Fundacion Tecnalia Research & Innovationspain); Thales Austria Gmbh (Austria); The Motor Insurance Repair Researchcentre (Uk); Toyota Motor Europe (Belgium); Nederlandse Organisatie Voor Toegepast Natuurwetenschappelijk Onderzoek Tno (Netherlands); Ttcontrol Gmbh (Austria); Tttech Computertechnik Ag (Austria); Technische Universiteit Eindhoven (Netherlands); Technische Universitat Darmstadt (Germany); Technische Universitaet Graz (Austria); Twt Gmbh Science & Innovation (Germany); University College Dublin, National University Of Ireland, Dublin (Ireland); Universidad De Las Palmas De Gran Canaria (Spain); Universita Degli Studi Di Modena E Reggio Emilia (Italy); Universidad Politecnica De Madrid (Spain); Valeo Autoklimatizace K.S. (Czech Republic); Valeo Comfort And Driving Assistance (France); Valeo Schalter Und Sensoren Gmbh (Germany); Kompetenzzentrum - Das Virtuelle Fahrzeug, Forschungsgesellschaft Mbh (Austria); Vires Simulationstechnologie Gmbh (Germany); Teknologian Tutkimuskeskus Vtt Oy (Finland); Tieto Finland Support Services Oy (Finland); Zilinska Univerzita V Ziline (Slovakia);

#### Inria contact: Axel Legay

The objective of ENABLE-S3 (http://www.enable-s3.eu) is to establish cost-efficient cross-domain virtual and semi-virtual V&V platforms and methods for ACPS. Advanced functional, safety and security test methods will be developed in order to significantly reduce the verification and validation time but preserve the validity of the tests for the requested high operation range. ENABLE-S3 aspires to substitute today's physical validation and verification efforts by virtual testing and verification, coverage-oriented test selection methods and standardization. ENABLE-S3 is use-case driven; these use cases represent relevant environments and scenarios. Each of the models, methods and tools integrated into the validation platform will be applied to at least one use case (under the guidance of the V&V methodology), where they will be validated (TRL 5) and their usability demonstrated (TRL6). Representative use cases and according applications provide the base for the requirements of methods and tools, as well as for the evaluation of automated systems and respective safety. This project is industry driven and has the objective of designing new technologies for autonomous transportation, including to secure them. Tamis tests its results on the case studies of the project.

Axel Legay and Jean-Louis Lanet are involved in this project. The project supports one postdoc in Tamis starting in 2017.

#### 9.3.1.5. SUCCESS

Title: SUCCESS: SecUre aCCESSibility for the internet of things

Program: CHIST-ERA 2015

Duration: 10/2016 - 10/2018

Coordinator: Middlesex University (UK)

Partners:

Middlesex University, School of Science and Technology (France); Inria (France); Université Grenoble Alpes, Verimag (FRANCE); University of TWENTE, (Netherlands)

#### Inria contact: Axel Legay

The SUCCESS project ...The core idea of SUCCESS is to use formal methods and verification tools with a proven track record to provide more transparency of security risks for people in given IoT scenarios. Our core scientific innovation will consist on the extension of well-known industry-strength methods Our technological innovation will provide adequate tools to address risk assessment and adaptivity within IoT in healthcare environments and an open source repository to foster future reuse, extension and progress in this area. Our project will validate the scientific and technological innovation through pilots, one of which will be in collaboration with a hospital and will allow all stakeholders (e.g. physicians, hospital technicians, patients and relatives) to enjoy a safer system capable to appropriately handle highly sensitive information on vulnerable people while making security and privacy risks understandable and secure solutions accessible.

Within SUCCESS, the contribution of the TAMIS team consists in a framework for analyzing the security of a given IOT system, and notably whether it resists to attack. Our approach is to build a high-level model of the system, including vulnerabilities, as well as an attacker. We represent the set of possible attacks using an attack tree. Finally, we evaluate the probability that an attack succeeds using Statistical Model Checking.

In the TAMIS team, Axel Legay, Delphine Beaulaton, Najah Ben-Saïd and Jean Quilbeuf are involved in this project.

#### 9.3.1.6. TeamPlay

Title: TeamPlay: Time, Energy and security Analysis for Multi/Many-core heterogeneous PLAtforms

Program: H2020

Duration: 01/2018 - 12/2020

Coordinator: Inria

Partners:

Absint Angewandte Informatik Gmbh (Germany), Institut National De Recherche en Informatique et Automatique (France), Secure-Ic Sas (France), Sky-Watch A/S (Danemark), Syddansk Universitet (Danemark), Systhmata Ypologistikis Orashs Irida Labs Ae (Greece), Technische Universität Hamburg-Harburg (Germany), Thales Alenia Space Espana (Spain), Universiteit Van Amsterdam (Netherlands), University Of Bristol (UK), University Of St Andrews (UK)

Inria contact: Olivier Zendra and Axel Legay

The TeamPlay (Time, Energy and security Analysis for Multi/Many-core heterogeneous PLAtforms) project federates 6 academic and 5 industrial partners and aims to develop new, formally-motivated, techniques that will allow execution time, energy usage, security, and other important non-functional properties of parallel software to be treated effectively, and as first- class citizens. We will build this into a toolbox for developing highly parallel software for low-energy systems, as required by

the internet of things, cyber-physical systems etc. The TeamPlay approach will allow programs to reflect directly on their own time, energy consumption, security, etc., as well as enabling the developer to reason about both the functional and the non-functional properties of their software at the source code level. Our success will ensure significant progress on a pressing problem of major industrial importance: how to effectively manage energy consumption for parallel systems while maintaining the right balance with other important software metrics, including time, security etc. The project brings together leading industrial and academic experts in parallelism, energy modeling/ transparency, worst-case execution time analysis, non-functional property analysis, compilation, security, and task coordination. Results will be evaluated using industrial use cases taken from the computer vision, satellites, flying drones, medical and cyber security domains. Within TeamPlay, Inria and TAMIS coordinate the whole project, while being also in charge of aspects related more specifically to security.

The permanent members of Tamis who are involved are Axel Legay, Olivier Zendra and Annelie Heuser.

# **10.** Dissemination

# **10.1. Promoting Scientific Activities**

## 10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- Axel Legay was General Chair for KimFest, an event organized for the 60th Birthday of Kim. G. Larsen
- Axel Legay was General Chair for the 1st ACM SAC Conference Track on Software-intensive Systems-of-Systems
- Jean-Louis Lanet was General Chair of Crisis 2017,
- Olivier Zendra was General co-Chair for ARCHI'17, the 9th Summer school on « Architecture des systèmes matériels et logiciels embarqués, et méthodes de conception associées »
- 10.1.1.2. Member of the Organizing Committees
  - Axel Legay coordinated KimFest, an event organized for the 60th Birthday of Kim. G. Larsen

#### 10.1.2. Scientific Events Selection

10.1.2.1. Member of Conference Steering Committees

- Axel Legay is a member of the Steering Committee of the Security summer school organized jointly by pre-GDR security and PEC (Pole d'Excellence Cyber).
- Jean-Louis Lanet has been member of the Steering Committee of Cardis 2017
- Olivier Zendra is a founder and a member of the Steering Committee of ICOOOLPS (International Workshop on Implementation, Compilation, Optimization of OO Languages, Programs and Systems)

#### 10.1.2.2. Chair of Conference Program Committees

- Axel Legay was Scientific chair of the 23rd International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)
- Axel Legay was the Scientific chair for the 11th International Conference on Risks and Security of Internet and Systems
- Axel Legay was the Scientific chair for the 17th International Conference on Application of Concurrency to System Design (ACSD 2017)

• Olivier Zendra was co-chair with Mario Wolzco of the Program Committee and the Organizing Committee of the 12th Workshop on Implementation, Compilation, Optimization of Object-Oriented Languages, Programs and Systems (ICOOOLPS 2017)

10.1.2.3. Member of the Conference Program Committees

- Axel Legay was a PC member of RV'17, ACSD'17, TACAS'17, CRISIS'17, CMSB'17, SETTA'17, FORMALIZE'17
- Fabrizio Biondi was a PC member of CRISIS'17, MCETECH'17, SAC'17
- Jean-Louis Lanet was PC member of Gramsec'17, Secitc'17, C2SI'17, Mcetech'17 and Afadl'17
- Olivier Zendra was PC member of ICOOOLPS'2017, ARCHI'17 and PEC 2017.

10.1.2.4. Reviewer

- Olivier Zendra was reviewer for MFCS.
- Fabrizio Biondi was a reviewer for CRISIS'17, ESORICS'17, KimFest, LATA'17, MFCS'17, RV'17, MCETECH'17

## 10.1.3. Journal

#### 10.1.3.1. Member of the Editorial Boards

- Axel Legay is a funder and member of the editorial board of "Foundations for Mastering Changes" journal.
- Annelie Heuser was PC Member/Editorial Board for IACR Transactions On Cryptographic Hardware And Embedded Systems
- 10.1.3.2. Reviewer Reviewing Activities
  - Axel Legay was reviewer for TCS, TSE, Information and Computation.
  - Annelie Heuser was a reviewer for Transactions on Information Forensics & Security, Journal of Cryptographic Engineering, Transactions on Embedded Computing Systems, IEEE Transactions on Very Large Scale Integration Systems
  - Jean-Louis was reviewer of Computer and Security journal

## 10.1.4. Invited Talks

- Axel Legay was an invited speaker for the 11th International Workshop on Reachability Problems.
- Axel Legay was invited speaker for the 43rd International Conference on Current Trends in Theory and Practice of Computer Science
- Fabrizio Biondi was invited speaker for The 12th International Conference on Risks and Security of Internet and Systems
- Florian Dold was invited to the "Re-Imagining Finance" workshop at Columbia Law School in New York City in September 2017.
- Annelie Heuser was invited to a panel discussion for Malicious Software and Hardware in Internet of Things (ACM International Conference on Computing Frontiers
- Jean-Louis Lanet was invited speaker for the INS3PECT workshop, the ROOTS conference, the Conference on Operational Planning, Technological Innovations and Mathematical Applications, the Journée AFSEC. 2017)

## 10.1.5. Scientific Expertise

- Axel Legay was an expert for the Wallonie Government.
- Axel Legay participated to the CR2 jury for Inria Nice Center as a member of Inria's evaluation committee.
- Olivier Zendra is a CIR expert for the MENESR.

- Olviier Zendra participated to the CR2 jury for Inria Paris Center as a member of Inria's evaluation committee.
- Olivier Zendra is a member of the editorial board and co-author of the "HiPEAC 2017 Vision" [47], as well as the HiPEAC 2019 Vision.

## 10.1.6. Research Administration

- Axel Legay is a member of Inria's evaluation committee.
- Axel Legay is the Representative for non-permanent staff committees (in charge of postdocs).
- Axel Legay is a member of "club équipe Française de la cyber sécurité"
- Axel Legay is the Britany Region representative in the ECSO organization
- Olivier Zendra is a member of Inria's evaluation committee.
- Olivier Zendra is a member of Inria's workgroup on Inria's social barometer.
- Olivier Zendra was a member of Inria's CNHSCT.
- Olivier Zendra was Head of Inria Nancy's IES Committee (formerly IST).

# 10.2. Teaching - Supervision - Juries

## 10.2.1. Teaching

- Master : Axel Legay, Introduction au Model Checking, 36, M2, Université de Bretagne Sud, France
- Master : Axel Legay, Introduction à l'analyse de risques, M2, Université de Bretagne Sud, France
- Licence : Nisrine Jafri , Programmation Java, L3, l'ISTIC, Université Rennes 1, France

## 10.2.2. Supervision

- PhD in progress : Kevin Bukasa, Démarrage sécurisé, 2015, Jean-Louis Lanet and Axel Legay
- PhD in progress : Mounir Chadli (Rennes 1), On Scheduling and SMC, December 2014, Axel Legay and Saddek Bensalem.
- PhD in progress : Olivier Descourbe, On Code Obfuscation, October 2016, Axel Legay and Fabrizio Biondi.
- PhD in progress : Mike Enescu, On Symbolic Execution for Malware Detection, October 2016, Axel Legay, Flavio Oquendo and Fabrizio Biondi. Terminated on October 2017.
- PhD in progress : Alexandre Gonsalvez, On Obfuscation via crypto primitives, April 2016, Axel Legay and Caroline Fontaine.
- PhD in progress : Nisrine Jafri (Rennes1), On fault Injection detection with MC of Binary code, December 2015, Axel Legay and Jean-Louis Lanet.
- PhD in progress : Razika Lounas, Validation des spécifications formelles de la mise à jour dynamique des applications Java Card, 2010, Mohamed Mezghiche and Jean-Louis Lanet
- PhD in progress: Martin Moreau (Rennes1); On the study of post-quantum cryptography mechanisms (provisory), Axel Legay, Annelie Heuser and Sylvain Guilley
- PhD in progress : Routa Moussaileb, From Data Signature to Behavior Analysis, 2017, Nora Cuppens and Jean-Louis Lanet
- PhD in progress : Tristan Ninet (Rennes 1), Vérification formelle d'une implémentation de la pile protocolaire IKEv2, December 2016, Axel Legay, Romaric Maillard and Olivier Zendra
- PhD in progress: Lamine Nouredine (Rennes1); Developing new packing detection techniques to stop malware propagation, November 2017, Axel Legay and Annelie Heuser.
- PhD in progress : Aurélien Palisse, Observabilité de codes hostiles, 2015, Jean-Louis Lanet

- PhD in progress: Emmanuel Tacheau (Rennes1); Analyse et détection de malwares au moyen de méthodes d'analyse symbolique, September 2017, Axel Legay, Fabrizio Biondi, Alain Fiocco.
- PhD in progress : Aurélien Trulla, Caractérisation de malware Android par suivi de flux d'information et nouvelles techniques d'évasion, 2016, Valerie Viet Triem Tong and Jean-Louis Lanet
- PhD in progress: Alexander Zhdanov (Rennes 1): Modular Automated Syntactic Signature Extraction (MASSE), December 2017, Axel Legay, Fabrizio Biondi, François Déchelle and Olivier Zendra.

## 10.2.3. Juries

- Axel Legay was a referee for the PhD defense of Xavier Devroye (University of Namur Belgium)
- Axel Legay was a referee for the PhD defense of Quentin Cappart (University of Louvain Belgium)
- Axel Legay was a referee for the PhD defense of Stefan Naujokat (University of Dortmund, Germany)

# **10.3.** Popularization

- Axel Legay participated to the "Forum Cyberstrategia" organized by the ministry of defense, September 2017
- Axel Legay participated to the "Inria Industry days" organized by Inria, October 2017
- Axel Legay participated to the "table ronde sur l'intelligence économique", Rennes November 2017
- Fabrizio Biondi participated to the "Forum International de la Cybersécurité", January 2017
- Fabrizio Biondi participated to the "Forum Cyberstrategia" organized by the ministry of defense, September 2017
- Fabrizio Biondi participated to the "Inria Industry days" organized by Inria, October 2017
- Fabrizio Biondi participated to the "European Cyber Week" organized by IRISA and Bretagne Development Innovation, November 2017

# 11. Bibliography

# **Publications of the year**

## **Doctoral Dissertations and Habilitation Theses**

- [1] Á. GARCÍA-RECUERO.Discouraging abusive behavior in privacy-preserving decentralized online social networks, Université Rennes 1, May 2017, https://tel.archives-ouvertes.fr/tel-01548658.
- [2] C. GROTHOFF. *The GNUnet System*, Université de Rennes 1, October 2017, Habilitation à diriger des recherches, https://hal.inria.fr/tel-01654244.

## **Articles in International Peer-Reviewed Journal**

[3] A. ARNOLD, M. BALEANI, A. FERRARI, M. MARAZZA, V. SENNI, A. LEGAY, J. QUILBEUF, C. ETZIEN. An Application of SMC to continuous validation of heterogeneous systems, in "EAI Endorsed Transactions on Industrial Networks and Intelligent Systems", February 2017, vol. 4, n<sup>o</sup> 10, p. 1-19 [DOI: 10.4108/EAI.1-2-2017.152154], https://hal.inria.fr/hal-01630523.

- [4] F. BIONDI, S. JOSSE, A. LEGAY, T. SIRVENT. Effectiveness of Synthesis in Concolic Deobfuscation, in "Computers and Security", September 2017, vol. 70, p. 500-515 [DOI : 10.1016/J.COSE.2017.07.006], https://hal.inria.fr/hal-01241356.
- [5] Y. CHEHERAZED, J.-L. LANET, M. MEZGHICHE, K. TAMINE. Machine Learning Techniques to Predict Sensitive Patterns to Fault Attack in the Java Card Application, in "Journal of Experimental and Theoretical Artificial Intelligence", 2017, https://hal.inria.fr/hal-01645392.
- [6] X. DEVROEY, G. PERROUIN, M. CORDY, H. SAMIH, A. LEGAY, P.-Y. SCHOBBENS, P. HEYMANS. Statistical Prioritization for Software Product Line Testing: an Experience Report, in "Software and Systems Modeling", February 2017, https://hal.inria.fr/hal-01642289.
- [7] F. DOLD, C. GROTHOFF. Byzantine set-union consensus using efficient set reconciliation, in "EURASIP Journal on Information Security", December 2017, vol. 2017, n<sup>o</sup> 1, p. 1-18 [DOI: 10.1186/s13635-017-0066-3], https://hal.inria.fr/hal-01657397.
- [8] J. DUCHÊNE, C. LE GUERNIC, E. ALATA, V. NICOMETTE, M. KAÂNICHE.State of the art of network protocol reverse engineering tools, in "Journal of Computer Virology and Hacking Techniques", 2017, 27 [DOI: 10.1007/s11416-016-0289-8], https://hal.inria.fr/hal-01496958.
- [9] N. EL MRABET, A. MRABET, R. LASHERMES, J.-B. RIGAUD, B. BOUALLEGUE, S. MESNAGER, M. MACHHOUT.A scalable and systolic architectures of montgomery modular multiplication for public key cryptosystems based on dsps, in "Journal Hardware and Systems Security", 2017, https://hal.archives-ouvertes.fr/hal-01579811.
- [10] A. HEUSER, S. PICEK, S. GUILLEY, N. MENTENS. Lightweight Ciphers and their Side-channel Resilience, in "IEEE Transactions on Computers", August 2017, p. 1-16 [DOI: 10.1109/TC.2017.2757921], https:// hal.archives-ouvertes.fr/hal-01629886.
- [11] J.-L. LANET, H. LE BOUDER, M. BENATTOU, A. LEGAY. When time meets test, in "International Journal of Information Security", 2017 [DOI: 10.1007/s10207-017-0371-3], https://hal.inria.fr/hal-01645395.
- [12] J.-L. LANET, A. MESBAH, M. MEZGHICHE. Reverse engineering a Java Card memory management algorithm, in "Computers & Security", May 2017, vol. 66, p. 97 - 114 [DOI: 10.1016/J.COSE.2017.01.005], https://hal.inria.fr/hal-01645393.
- [13] R. LOUNAS, J.-L. LANET, M. MEZGHICHE. A Formal Verification of Dynamic Updating in a Java-based embedded System, in "International Journal of Critical Computer-Based Systems", 2017, https://hal.inria.fr/ hal-01645401.
- [14] B. NICOLAS, C. CARLET, S. GUILLEY, A. HEUSER, E. PROUFF, O. RIOUL. Stochastic Collision Attack, in "IEEE Transactions on Information Forensics and Security", April 2017, vol. 12, n<sup>o</sup> 9, p. 2090 - 2104 [DOI: 10.1109/TIFS.2017.2697401], https://hal.archives-ouvertes.fr/hal-01629880.
- [15] B. NICOLAS, S. GUILLEY, A. HEUSER, M. DAMIEN, O. RIOUL. Optimal side-channel attacks for multivariate leakages and multiple models, in "Journal of Cryptographic Engineering", August 2017, vol. 7, n<sup>o</sup> 4, p. 331–341 [DOI: 10.1007/s13389-017-0170-9], https://hal.archives-ouvertes.fr/hal-01629885.
- [16] S. PICEK, A. HEUSER, S. GUILLEY. Template attack versus Bayes classifier, in "Journal of Cryptographic Engineering", September 2017, vol. 7, n<sup>o</sup> 4, p. 343–351 [DOI: 10.1007/s13389-017-0172-7], https://hal. archives-ouvertes.fr/hal-01629884.
- [17] È. DE CHÈRISEY, S. GUILLEY, A. HEUSER, O. RIOUL.On the optimality and practicability of mutual information analysis in some scenarios, in "Cryptography and Communications - Discrete Structures, Boolean Functions and Sequences", June 2017, https://hal.archives-ouvertes.fr/hal-01645127.

#### **International Conferences with Proceedings**

- [18] F. BIONDI, M. CHADLI, T. GIVEN-WILSON, A. LEGAY.*Information Leakage as a Scheduling Resource*, in "International Workshop on Formal Methods for Industrial Critical Systems and Automated Verification of Critical Systems", Turin, Italy, September 2017, https://hal.inria.fr/hal-01382052.
- [19] F. BIONDI, F. DÉCHELLE, A. LEGAY.MASSE: Modular Automated Syntactic Signature Extraction, in "ISSRE 2017 - The 28th International Symposium on Software Reliability Engineering - IEEE", Toulouse, France, IEEE, October 2017, p. 1-2, https://hal.inria.fr/hal-01629035.
- [20] F. BIONDI, Y. KAWAMOTO, A. LEGAY, L.-M. TRAONOUEZ.*HyLeak: Hybrid Analysis Tool for Information Leakage*, in "ATVA 2017 Fifteenth International Symposium on Automated Technology for Verification and Analysis", Pune, India, October 2017, 14, https://hal.inria.fr/hal-01546817.
- [21] A. BKAKRIA, M. GRAA, N. CUPPENS-BOULAHIA, F. CUPPENS, J.-L. LANET. Real-time detection and reaction to Activity hijacking attacks in Android smartphones, in "PST 2017, Privacy, Security, and Trust -15th International Conference", Calgary, Canada, August 2017, https://hal.inria.fr/hal-01645399.
- [22] Q. CAPPART, C. LIMBRÉE, P. SCHAUS, J. QUILBEUF, L.-M. TRAONOUEZ, A. LEGAY. Verification of Interlocking Systems Using Statistical Model Checking, in "18th IEEE International Symposium on High Assurance Systems Engineering (HASE)", Singapore, Singapore, 18th IEEE International Symposium on High Assurance Systems Engineering (HASE), January 2017, p. 61 - 68 [DOI: 10.1109/HASE.2017.10], https://hal.archives-ouvertes.fr/hal-01591338.
- [23] C. CARLET, A. HEUSER, S. PICEK.*Trade-Offs for S-Boxes: Cryptographic Properties and Side-Channel Resilience*, in "ACNS 2017 International Conference on Applied Cryptography and Network Security", Kanazawa, Japan, LNCS, Springer, July 2017, vol. 10355, p. 393-414 [DOI : 10.1007/978-3-319-61204-1\_20], https://hal.archives-ouvertes.fr/hal-01629879.
- [24] M. CONTINI, G. DI NATALE, A. HEUSER, T. POPPELMANN, N. MENTENS. Do we need a holistic approach for the design of secure IoT systems?, in "Computing Frontiers Conference", Siena, Italy, May 2017, https:// hal.archives-ouvertes.fr/hal-01628683.
- [25] X. DEVROEY, G. PERROUIN, M. PAPADAKIS, A. LEGAY, P.-Y. SCHOBBENS, P. HEYMANS. Automata Language Equivalence vs. Simulations for Model-based Mutant Equivalence: An Empirical Evaluation, in "ICST 2017 - International Conference on Software Testing, Verification and Validation", tokyo, Japan, March 2017, https://hal.inria.fr/hal-01640101.
- [26] T. GIVEN-WILSON, N. JAFRI, J.-L. LANET, A. LEGAY. An Automated Formal Process for Detecting Fault Injection Vulnerabilities in Binaries and Case Study on PRESENT, in

"2017 IEEE Trustcom/BigDataSE/ICESS", Sydney, Australia, August 2017, p. 293 - 300 [DOI: 10.1109/TRUSTCOM/BIGDATASE/ICESS.2017.250], https://hal.inria.fr/hal-01629098.

- [27] T. GIVEN-WILSON, A. LEGAY, S. SEDWARDS. Information Security, Privacy, and Trust in Social Robotic Assistants for Older Adults, in "HAS 2017 - International Conference on Human Aspects of Information Security, Privacy, and Trust", Vancouver, Canada, T. TRYFONAS (editor), LNCS, Springer, July 2017, vol. 10292, p. 90-109 [DOI: 10.1007/978-3-319-58460-7\_7], https://hal.inria.fr/hal-01629094.
- [28] P. GJØL JENSEN, K. G. LARSEN, A. LEGAY, D. B. POULSEN. Quantitative Evaluation of Attack Defense Trees using Stochastic Timed Automata, in "GraMSec 2017 - The Fourth International Workshop on Graphical Models for Security", Santa Barbara, United States, August 2017, https://hal.inria.fr/hal-01640091.
- [29] M. GRAA, N. CUPPENS-BOULAHIA, F. CUPPENS, J.-L. LANET, R. MOUSSAILEB. Detection of Side Channel Attacks Based on Data Tainting in Android Systems, in "SEC 2017 - 32th IFIP International Conference on ICT Systems Security and Privacy Protection", Rome, Italy, S. D. C. DI VIMERCATI, F. MARTINELLI (editors), ICT Systems Security and Privacy Protection, Springer International Publishing, May 2017, vol. AICT-502, p. 205-218, Part 4: Operating System and Firmware Security [DOI : 10.1007/978-3-319-58469-0\_14], https://hal.inria.fr/hal-01648994.
- [30] S. GUILLEY, A. HEUSER, O. RIOUL. Codes for Side-Channel Attacks and Protections, in "C2SI 2017 -International Conference on Codes, Cryptology, and Information Security", Rabat, Morocco, LNCS, April 2017, vol. 10194, p. 35-55 [DOI : 10.1007/978-3-319-55589-8\_3], https://hal.archives-ouvertes.fr/hal-01629876.
- [31] B. KEVIN, R. LASHERMES, J.-L. LANET, H. LE BOUDER, A. LEGAY.*How TrustZone could be bypassed: Side-Channel Attacks on a modern System-on-Chip*, in "Wistp'17, International Conference on Information Security Theory and Practice", Heraklion, Greece, September 2017, https://hal.inria.fr/hal-01645398.
- [32] A. KUNNAPPILLY, A. LEGAY, T. MARGARIA, C. SECELEANU, B. STEFFEN, L.-M. TRAONOUEZ. Analyzing Ambient Assisted Living Solutions: A Research Perspective, in "12th International Conference on Desig &Technology of Integrated Systems In Nanoscale Era (DTIS)", Palma de Mallorca, Spain, 12th International Conference on Desig &Technology of Integrated Systems In Nanoscale Era (DTIS), April 2017, p. 1 - 7 [DOI: 10.1109/DTIS.2017.7930168], https://hal.archives-ouvertes.fr/hal-01591347.
- [33] J.-L. LANET. Experimenting similarity-based hijacking attacks detection and response in Android Systems, in "ICISS, International Conference on Information Systems Security", Bombay, India, December 2017, https:// hal.inria.fr/hal-01645400.
- [34] J.-L. LANET, A. PALISSE. When data reveals ransomware activity, in "4th International Conference on Operational Planning, Technological Innovations and Mathematical Applications,", Athen, Greece, May 2017, https://hal.inria.fr/hal-01645396.
- [35] C. LE GUERNIC.Toward a Sound Analysis of Guarded LTI Loops with Inputs by Abstract Acceleration (extended version), in "Static Analysis Symposium", New York, United States, Lecture Notes in Computer Science, August 2017, vol. 10422, https://hal.inria.fr/hal-01550767.
- [36] G. LI, P. M. JENSEN, K. G. LARSEN, A. LEGAY, D. B. POULSEN. Practical Controller Synthesis for MTL0, ∞, in "International SPIN Symposium on Model Checking of Software", Santa barbara, United States, July 2017, https://hal.inria.fr/hal-01642162.

- [37] G. PERROUIN, P. HEYMANS, A. LEGAY, X. DEVROEY, M. CORDY, P.-Y. SCHOBBENS. On Featured Transition Systems, in "SOFSEM 2017 - 43rd International Conference on Current Trends in Theory and Practice of Informatics", Limerick, Ireland, January 2017, https://hal.inria.fr/hal-01640267.
- [38] S. PICEK, A. HEUSER, A. JOVIC, A. LEGAY. Climbing Down the Hierarchy: Hierarchical Classification for Machine Learning Side-Channel Attacks, in "AFRICACRYPT 2017 - International Conference on Cryptology in Africa", Dakar, Senegal, LNCS, Springer, May 2017, vol. 10239, p. 61-78 [DOI : 10.1007/978-3-319-57339-7\_4], https://hal.archives-ouvertes.fr/hal-01629878.

#### **Conferences without Proceedings**

- [39] S. GUILLEY, A. HEUSER, M. TANG, O. RIOUL.Stochastic Side-Channel Leakage Analysis via Orthonormal Decomposition, in "SecITC 2017", Bucharest, Romania, June 2017, https://hal.archives-ouvertes.fr/hal-01628679.
- [40] J.-L. LANET. *Formal Methods and the Dark Side of the Force*, in "Journée Approches Formelles des Systèmes Embarqués Communicants", Paris, France, June 2017, https://hal.inria.fr/hal-01645403.
- [41] J.-L. LANET, A. MESBAH. The Express Laundry from black box to white box, in "Workshop InS3PECT: Ingénierie Système de Services Sécurisés Pour objEts ConnecTé", Nice, France, December 2017, https://hal. inria.fr/hal-01645404.
- [42] A. LEGAY, L.-M. TRAONOUEZ. Plasma Lab Statistical Model Checker: Architecture, Usage and Extension, in "SOFSEM 2017 - 43rd International Conference on Current Trends in Theory and Practice of Computer Science", Limerick, Ireland, January 2017, https://hal.archives-ouvertes.fr/hal-01613581.
- [43] A. MESBAH, J.-L. LANET, M. MEZGHICHE. Reverse Engineering a Code without the Code, in "1st Reversing and Offensive-oriented Trends Symposium 2017", Vienna, Austria, Sergey Bratus (Dartmouth College), November 2017, https://hal.inria.fr/hal-01591926.
- [44] A. MESBAH, M. MEZGHICHE, J.-L. LANET. Persistant Fault Injection Attack, From White-box to Black-box, in "The 5th International Conference on Electrical Engineering - ICEE 2017", Boumedrès, Algeria, October 2017, https://hal.inria.fr/hal-01591931.
- [45] A. PALISSE, A. DURAND, H. LE BOUDER, C. LE GUERNIC, J.-L. LANET. Data Aware Defense (DaD): Towards a Generic and Practical Ransomware Countermeasure, in "NordSec 2017", Tartu, Estonia, November 2017, https://hal.inria.fr/hal-01591939.
- [46] S. PICEK, A. HEUSER, A. JOVIC, S. LUDWIG, S. GUILLEY, D. JAKOBOVIĆ, N. MENTENS. Side-channel analysis and machine learning: A practical perspective, in "International Joint Conference on Neural Networks (IJCNN)", Anchorage, United States, May 2017, https://hal.archives-ouvertes.fr/hal-01628681.

#### **Books or Proceedings Editing**

- [47] M. DURANTON, K. DE BOSSCHERE, C. GAMRAT, J. MAEBE, H. MUNK, O. ZENDRA (editors). *The HiPEAC Vision 2017*, HiPEAC network of excellence, January 2017, 138, https://hal.inria.fr/hal-01491758.
- [48] F. OQUENDO, K. DRIRA, A. LEGAY, T. BATISTA (editors). Proceedings of the 1st ACM SAC Conference Track on Software-intensive Systems-of-Systems (SiSoS 2017): 32nd ACM SIGAPP Symposium On Applied Computing, ACM, Marrakesh, Morocco, April 2017, https://hal.archives-ouvertes.fr/hal-01445350.

#### **Scientific Popularization**

[49] A. PALISSE, J. JONGWANE. Quelles solutions pour se protéger des logiciels de rançon ?, in "Interstices", October 2017, https://hal.inria.fr/hal-01688779.

#### **Other Publications**

- [50] N. BEN SAID, T. ABDELLATIF, M. BOZGA, S. BEN SALEM, A. LEGAY. Orchestration for Secure Multiparty Communications in Web-Services ☆, November 2017, working paper or preprint, https://hal.inria.fr/ hal-01629427.
- [51] N. BEN SAID, F. BIONDI, V. BONTCHEV, O. DECOURBE, T. GIVEN-WILSON, A. LEGAY, J. QUIL-BEUF. Detection of Mirai by Syntactic and Semantic Analysis, November 2017, working paper or preprint, https://hal.inria.fr/hal-01629040.
- [52] F. BIONDI, M. A. ENESCU, A. HEUSER, A. LEGAY, K. S. MEEL, J. QUILBEUF. Scalable Approximation of Quantitative Information Flow in Programs, November 2017, working paper or preprint, https://hal.inria.fr/ hal-01629131.
- [53] F. BIONDI, T. GIVEN-WILSON, A. LEGAY. *Universal Optimality of Apollonian Cell Encoders*, October 2017, working paper or preprint, https://hal.inria.fr/hal-01571226.
- [54] F. BIONDI, Y. KAWAMOTO, A. LEGAY, L.-M. TRAONOUEZ. Hybrid Statistical Estimation of Mutual Information and its Application to Information Flow, November 2017, working paper or preprint, https:// hal.inria.fr/hal-01629033.
- [55] M. CHADLI, J. H. KIM, K. G. LARSEN, A. LEGAY, S. NAUJOKAT, B. STEFFEN, L.-M. TRAONOUEZ. *High-level Frameworks for the Specification and Verification of Scheduling Problems*, October 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01613576.
- [56] T. GIVEN-WILSON, A. HEUSER, N. JAFRI, J.-L. LANET, A. LEGAY. An Automated and Scalable Formal Process for Detecting Fault Injection Vulnerabilities in Binaries, November 2017, working paper or preprint, https://hal.inria.fr/hal-01629135.
- [57] T. GIVEN-WILSON, N. JAFRI, J.-L. LANET, A. LEGAY. An Automated Formal Process for Detecting Fault Injection Vulnerabilities in Binaries and Case Study on PRESENT – Extended Version, April 2017, working paper or preprint, https://hal.inria.fr/hal-01400283.
- [58] T. GIVEN-WILSON, A. LEGAY, S. SEDWARDS, O. ZENDRA. Modelling of Machine-Aided Human Group Motion, November 2017, working paper or preprint, https://hal.inria.fr/hal-01629137.
- [59] A. HEUSER, S. PICEK, A. JOVIC, A. LEGAY. On the Relevance of Feature Selection for Profiled Side-channel Attacks, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01645128.
- [60] A. HEUSER, S. PICEK, A. LEGAY, K. KNEZEVIC. *Profiled SCA with a New Twist: Semi-supervised Learning*, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01645130.
- [61] L.-M. TRAONOUEZ, A. LEGAY, D. NOWOTKA, D. B. POULSEN. *Statistical Model Checking of LLVM Code*, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01640097.

#### **References in notes**

- [62] M. CHADLI, J. H. KIM, A. LEGAY, L.-M. TRAONOUEZ, S. NAUJOKAT, B. STEFFEN, K. G. LARSEN.A Model-Based Framework for the Specification and Analysis of Hierarchical Scheduling Systems, in "FMICS-AVoCS", Pise, Italy, Critical Systems: Formal Methods and Automated Verification, Springer, September 2016, vol. 9933, p. 133 - 141 [DOI: 10.1007/978-3-319-45943-1\_9], https://hal.archives-ouvertes.fr/hal-01241681.
- [63] Á. GARCÍA-RECUERO, J. BURDGES, C. GROTHOFF. Privacy-Preserving Abuse Detection in Future Decentralised Online Social Networks, in "11th International ESORICS Workshop in Data Privacy Management, DPM 2016", Heraklion, Crete, Greece, G. LIVRAGA, V. TORRA, A. ALDINI, F. MARTINELLI, N. SURI (editors), Springer Lecture Notes in Computer Science (LNCS) series, Springer, September 2016, vol. 9963, p. 78-93 [DOI: 10.1007/978-3-319-47072-6\_6], https://hal.inria.fr/hal-01355951.
- [64] A. HEUSER, S. PICEK, S. GUILLEY, N. MENTENS. Side-channel Analysis of Lightweight Ciphers: Current Status and Future Directions, October 2016, Lightweight Cryptography Workshop 2016, https://hal.inria.fr/ hal-01407264.
- [65] A. HEUSER, S. PICEK, S. GUILLEY, N. MENTENS. Side-channel Analysis of Lightweight Ciphers: Does Lightweight Equal Easy?, in "RFIDSec 2016: 12th Workshop on RFID and IoT Security", Hong Kong, Hong Kong SAR China, November 2016, https://hal.inria.fr/hal-01402238.
- [66] A. SAVARY, M. FRAPPIER, M. LEUSCHEL, J. LANET.*Model-Based Robustness Testing in Event-B Using Mutation*, in "Software Engineering and Formal Methods 13th International Conference, SEFM 2015, York, UK, September 7-11, 2015. Proceedings", R. CALINESCU, B. RUMPE (editors), Lecture Notes in Computer Science, Springer, 2015, vol. 9276, p. 132–147, http://dx.doi.org/10.1007/978-3-319-22969-0\_10.
- [67] N. H. WALFIELD, J. L. GRIFFIN, C. GROTHOFF.A Quantitative Analysis of Cell Tower Trace Data for Understanding Human Mobility and Mobile Networks, in "6th International Workshop on Mobile Entity Localization, Tracking and Analysis (MELT)", San Francisco, United States, October 2016 [DOI: 10.1145/1235], https://hal.inria.fr/hal-01378622.

# **Project-Team TEA**

## Time, Events and Architectures

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Embedded and Real-time Systems

## **Table of contents**

1.	Personnel	. 981
2.	Overall Objectives	. 982
	2.1. Introduction	982
	2.2. Context	982
	2.3. Motivations	983
	2.4. Challenges	983
3.	Research Program	. 984
	3.1. Previous Works	984
	3.2. Modeling Times	985
	3.3. Modeling Architectures	985
	3.4. Scheduling Theory	986
	3.5. Virtual Prototyping	987
4.	Application Domains	. 987
	4.1. Automotive and Avionics	987
	4.2. Factory Automation	988
5.	Highlights of the Year	. 989
6.	New Software and Platforms	. 989
	6.1. ADFG	989
	6.2. POLYCHRONY	989
	6.3. Polychrony AADL2SIGNAL	990
	6.4. POP	990
7	0.5. Sigali	991
7.	7.1 ADEC: Affine data flow graphs schedular synthesis	. 991
	7.1. ADFO. Annue data-now graphs scheduler synthesis	991
	7.2. Format Semantics of Denavior Specifications in the Arcintecture Analysis and Design L	002
	7.3 New trends and developments in Polychrony	003
	7.4 Modular verification of cyber-physical systems using contract theory	994
	7.5 Parametric verification of time synchronization protocols	995
	7.6 Modular analysis and verification of system libraries	995
8.	Bilateral Contracts and Grants with Industry	. 996
9.	Partnerships and Cooperations	.996
	9.1. National Initiatives	996
	9.1.1. ANR	996
	9.1.2. PAI	996
	9.2. International Initiatives	996
	9.2.1. Inria International Labs	996
	9.2.2. Inria Associate Teams	997
	9.2.3. Inria International Partners	998
	9.3. International Research Visitors	998
	9.3.1. Visits of International Scientists	998
	9.3.2. Visits to International Teams	998
10.	Dissemination	. 999
	10.1. Promoting Scientific Activities	999
	10.1.1. Scientific events organisation	999
	10.1.1.1. General chair, scientific chair	999
	10.1.1.2. Member of the organizing committees	999
	10.1.2. Scientific Events Selection	999
	10.1.3. Journal	999

10.1.3.1. Member of the Editorial Boards	999
10.1.3.2. Reviews	999
10.1.4. Invited Talks	999
10.1.5. Invited Talks	999
10.1.6. Scientific Expertise	999
10.2. Teaching - Supervision - Juries	999
10.2.1. Teaching	999
10.2.2. Supervision	999
10.2.3. Juries	1000
11. Bibliography	<b>1000</b>

### **Project-Team TEA**

*Creation of the Team: 2014 January 01, updated into Project-Team: 2015 January 01* **Keywords:** 

#### **Computer Science and Digital Science:**

- A1.2.7. Cyber-physical systems
- A1.5.2. Communicating systems
- A2.1.1. Semantics of programming languages
- A2.1.3. Functional programming
- A2.1.6. Concurrent programming
- A2.1.8. Synchronous languages
- A2.1.10. Domain-specific languages
- A2.2.1. Static analysis
- A2.2.4. Parallel architectures
- A2.3. Embedded and cyber-physical systems
- A2.3.1. Embedded systems
- A2.3.2. Cyber-physical systems
- A2.3.3. Real-time systems
- A2.4. Verification, reliability, certification
- A2.4.1. Analysis
- A2.4.2. Model-checking
- A2.4.3. Proofs
- A2.5. Software engineering
- A2.5.1. Software Architecture & Design
- A2.5.2. Component-based Design
- A4.4. Security of equipment and software
- A4.5. Formal methods for security
- A7.2. Logic in Computer Science
- A7.2.3. Interactive Theorem Proving
- A7.3. Computational models
- A8.1. Discrete mathematics, combinatorics
- A8.4. Computer Algebra

#### **Other Research Topics and Application Domains:**

- B5.1. Factory of the future
- B6.1.1. Software engineering
- B6.4. Internet of things
- B6.6. Embedded systems

## 1. Personnel

#### **Research Scientists**

Jean-Pierre Talpin [Team leader, Inria, Senior Researcher, HDR]

Thierry Gautier [Inria, Researcher] Rajesh Kumar Gupta [Inria, Advanced Research Position, from Aug. 2017] Vania Joloboff [Inria, Researcher, until Sep. 2017]

#### **Technical Staff**

Loïc Besnard [CNRS SED, Senior Research Engineer, seconded at 60%] Hai Nam Tran [Inria] Alexandre Honorat [Inria, until Sep. 2017]

#### **PhD Students**

Simon Lunel [Mitsubishi Electric] Jean-Joseph Marty [Inria, from Oct. 2017] Liangcong Zhang [China Scholarship Council, from Oct. 2017]

#### Visiting Scientists

Alexandre Honorat [IETR, from Oct. 2017] Deian Stefan [UC San Diego, Sep. 2016]

Administrative Assistant

Armelle Mozziconacci [CNRS]

## 2. Overall Objectives

#### 2.1. Introduction

An embedded architecture is an artifact of heterogeneous constituents and at the crossing of several design viewpoints: software, embedded in hardware, interfaced with the physical world. Time takes different forms when observed from each of these viewpoints: continuous or discrete, event-based or time-triggered. Unfortunately, modeling and programming formalisms that represent software, hardware and physics significantly alter this perception of time. Moreover, time reasoning in system design is usually isolated to a specific design problem: simulation, profiling, performance, scheduling, parallelization, simulation. The aim of project-team TEA is to define conceptually unified frameworks for reasoning on composition and integration in cyber-physical system design, and to put this reasoning to practice by revisiting analysis and synthesis issues in real-time system design with soundness and compositionality gained from formalization.

#### 2.2. Context

In the construction of complex systems, information technology (IT) has become a central force of revolutionary changes, driven by the exponential increase of computational power. In the field of telecommunication, IT provides the necessary basis for systems of networked distributed applications. In the field of control engineering, IT provides the necessary basis for embedded control applications. The combination of telecommunication and embedded systems into networked embedded systems opens up a new range of systems, capable of providing more intelligent functionality thank to information and communication (ICT). Networked embedded systems have revolutionized several application domains: energy networks, industrial automation and transport systems.

20th-century science and technology brought us effective methods and tools for designing both computational and physical systems. But the design of cyber-physical systems (CPS) is much more than the union of those two fields. Traditionally, information scientists only have a hazy notion of requirements imposed by the physical environment of computers. Similarly, mechanical, civil, and chemical engineers view computers strictly as devices executing algorithms. To the extent we have designed CPS, we have done so in an ad hoc, on-off manner that is not repeatable. A new science of CPS design will allow us to create new machines with complex dynamics and high reliability, to apply its principles to new industries and applications in a reliable and economically efficient way. Progress requires nothing less than the construction of a new science and technology foundation for CPS that is simultaneously physical and computational.

#### 2.3. Motivations

Beyond the buzzword, a CPS is an ubiquitous object of our everyday life. CPSs have evolved from individual independent units (e.g an ABS brake) to more and more integrated networks of units, which may be aggregated into larger components or sub-systems. For example, a transportation monitoring network aggregates monitored stations and trains through a large scale distributed system with relatively high latency. Each individual train is being controlled by a train control network, each car in the train has its own real-time bus to control embedded devices. More and more, CPSs are mixing real-time low latency technology with higher latency distributed computing technology.

In the past 15 years, CPS development has moved towards Model Driven Engineering (MDE). With MDE methodology, first all requirements are gathered together with use cases, then a model of the system is built (sometimes several models) that satisfy the requirements. There are several modeling formalisms that have appeared in the past ten years with more or less success. The most successful are the *executable* models<sup>000</sup>, i.e., models that can be simulated, exercised, tested and validated. This approach can be used for both software and hardware.

A common feature found in CPSs is the ever presence of concurrency and parallelism in models. Large systems are increasingly mixing both types of concurrency. They are structured hierarchically and comprise multiple synchronous devices connected by buses or networks that communicate asynchronously. This led to the advent of so-called GALS (Globally Asynchronous, Locally Synchronous) models, or PALS (Physically Asynchronous), Logically Synchronous) systems, where reactive synchronous objects are communicating asynchronously. Still, these infrastructures, together with their programming models, share some fundamental concerns: parallelism and concurrency synchronization, determinism and functional correctness, scheduling optimality and calculation time predictability.

Additionally, CPSs monitor and control real-world processes, the dynamics of which are usually governed by physical laws. These laws are expressed by physicists as mathematical equations and formulas. Discrete CPS models cannot ignore these dynamics, but whereas the equations express the continuous behavior usually using real numbers (irrational) variables, the models usually have to work with discrete time and approximate floating point variables.

#### 2.4. Challenges

A cyber-physical, or reactive, or embedded system is the integration of heterogeneous components originating from several design viewpoints: reactive software, some of which is embedded in hardware, interfaced with the physical environment through mechanical parts. Time takes different forms when observed from each of these viewpoints: it is discrete and event-based in software, discrete and time-triggered in hardware, continuous in mechanics or physics. Design of CPS often benefits from concepts of multiform and logical time(s) for their natural description. High-level formalisms used to model software, hardware and physics additionally alter this perception of time quite significantly.

In model-based system design, time is usually abstracted to serve the purpose of one of many design tasks: verification, simulation, profiling, performance analysis, scheduling analysis, parallelization, distribution, or virtual prototyping. For example in non-real-time commodity software, timing abstraction such as number of instructions and algorithmic complexity is sufficient: software will run the same on different machines, except slower or faster. Alternatively, in cyber-physical systems, multiple recurring instances of meaningful events may create as many dedicated logical clocks, on which to ground modeling and design practices.

Time abstraction increases efficiency in event-driven simulation or execution (i.e SystemC simulation models try to abstract time, from cycle-accurate to approximate-time, and to loosely-time), while attempting to retain functionality, but without any actual guarantee of valid accuracy (responsibility is left to the model designer). Functional determinism (a.k.a. conflict-freeness in Petri Nets, monotonicity in Kahn PNs, confluence in

<sup>&</sup>lt;sup>0</sup>Matlab/Simulink, https://fr.mathworks.com/products/simulink.html

<sup>&</sup>lt;sup>0</sup>*Ptolemy*, http://ptolemy.eecs.berkeley.edu

<sup>&</sup>lt;sup>0</sup>SysML, http://www.uml-sysml.org

Milner's CCS, latency-insensitivity and elasticity in circuit design) allows for reducing to some amount the problem to that of many schedules of a single self-timed behavior, and time in many systems studies is partitioned into models of computation and communication (MoCCs). Multiple, multiform time(s) raises the question of combination, abstraction or refinement between distinct time bases. The question of combining continuous time with discrete logical time calls for proper discretization in simulation and implementation. While timed reasoning takes multiple forms, there is no unified foundation to reasoning about multi-form time in system design.

The objective of project-team TEA is henceforth to define formal models for timed quantitative reasoning, composition, and integration in embedded system design. Formal time models and calculi should allow us to revisit common domain problems in real-time system design, such as time predictability and determinism, memory resources predictability, real-time scheduling, mixed-criticality and power management; yet from the perspective gained from inter-domain timed and quantitative abstraction or refinement relations. A regained focus on fundamentals will allow to deliver better tooled methodologies for virtual prototyping and integration of embedded architectures.

## **3. Research Program**

#### 3.1. Previous Works

The challenges of team TEA support the claim that sound Cyber-Physical System design (including embedded, reactive, and concurrent systems altogether) should consider multi-form time models as a central aspect. In this aim, architectural specifications found in software engineering are a natural focal point to start from. Architecture descriptions organize a system model into manageable components, establish clear interfaces between them, collect domain-specific constraints and properties to help correct integration of components during system design. The definition of a formal design methodology to support heterogeneous or multi-form models of time in architecture descriptions demands the elaboration of sound mathematical foundations and the development of formal calculi and methods to instrument them. This constitutes the research program of team TEA.

System design based on the "synchronous paradigm" has focused the attention of many academic and industrial actors on abstracting non-functional implementation details from system design. This elegant design abstraction focuses on the logic of interaction in reactive programs rather than their timed behavior, allowing to secure functional correctness while remaining an intuitive programming model for embedded systems. Yet, it corresponds to embedded technologies of single cores and synchronous buses from the 90s, and may hardly cover the semantic diversity of distribution, parallelism, heterogeneity, of cyber-physical systems found in 21st century Internet-connected, true-time $^{TM}$ -synchronized clouds, of tomorrow's grids.

By contrast with a synchronous hypothesis yet from the same era, the polychronous MoCC implemented in the data-flow specification language Signal, available in the Eclipse project POP <sup>0</sup> and in the CCSL standard. <sup>0</sup>, are inherently capable of describing multi-clock abstractions of GALS systems. The POP and TimeSquare projects provide tooled infrastructures to refine high-level specifications into real-time streaming applications or locally synchronous and globally asynchronous systems, through a series of model analysis, verification, and synthesis services. These tool-supported refinement and transformation techniques can assist the system engineer from the earliest design stages of requirement specification to the latest stages of synthesis, scheduling and deployment. These characteristics make polychrony much closer to the required semantic for compositional, refinement-based, architecture-driven, system design.

While polychrony was a step ahead of the traditional synchronous hypothesis, CCSL is a leap forward from synchrony and polychrony. The essence of CCSL is "multi-form time" toward addressing all of the domain-specific physical, electronic and logical aspects of cyber-physical system design.

<sup>&</sup>lt;sup>0</sup>Polychrony on Polarsys, https://www.polarsys.org/projects/polarsys.pop

<sup>&</sup>lt;sup>0</sup>Clock Constraints in UML/MARTE CCSL. C. André, F. Mallet. RR-6540. Inria, 2008. http://hal.inria.fr/inria-00280941

#### **3.2. Modeling Times**

To make a sense and eventually formalize the semantics of time in system design, we should most certainly rely on algebraic representations of time found in previous works and introduce the paradigm of "time systems" (type systems to represent time) in a way reminiscent to CCSL. Just as a type system abstracts data carried along operations in a program, a time system abstracts the causal interaction of that program module or hardware element with its environment, its pre and post conditions, its assumptions and guarantees, either logical or numerical, discrete or continuous. Some fundamental concepts of the time systems we envision are present in the clock calculi found in data-flow synchronous languages like Signal or Lustre, yet bound to a particular model of concurrency, hence time.

In particular, the principle of refinement type systems <sup>0</sup>, is to associate information (data-types) inferred from programs and models with properties pertaining, for instance, to the algebraic domain on their value, or any algebraic property related to its computation: effect, memory usage, pre-post condition, value-range, cost, speed, time, temporal logic <sup>0</sup>. Being grounded on type and domain theories, a time system should naturally be equipped with program analysis techniques based on type inference (for data-type inference) or abstract interpretation (for program properties inference) to help establish formal relations between heterogeneous component "types". Just as a time calculus may formally abstract timed concurrent behaviors of system components, timed relations (abstraction and refinement) represent interaction among components.

Scalability and compositionality requires the use of assume-guarantee reasoning to represent them, and to facilitate composition by behavioral sub-typing, in the spirit of the (static) contract-based formalism proposed by Passerone et al. <sup>0</sup>. Verification problems encompassing heterogeneously timed specifications are common and of great variety: checking correctness between abstract and concrete time models relates to desynchronisation (from synchrony to asynchrony) and scheduling analysis (from synchrony to hardware). More generally, they can be perceived from heterogeneous timing viewpoints (e.g. mapping a synchronous-time software on a real-time middle-ware or hardware).

This perspective demands capabilities not only to inject time models one into the other (by abstract interpretation, using refinement calculi), to compare time abstractions one another (using simulation, refinement, bi-simulation, equivalence relations) but also to prove more specific properties (synchronization, determinism, endochrony). All this formalization effort will allow to effectively perform the tooled validation of common cross-domain properties (e.g. cost v.s. power v.s. performance v.s. software mapping) and tackle equally common yet though case studies such as these linking battery capacity, to on-board CPU performance, to static software schedulability, to logical software correctness and plant controllability: the choice of the right sampling period across the system components.

#### **3.3. Modeling Architectures**

To address the formalization of such cross-domain case studies, modeling the architecture formally plays an essential role. An architectural model represents components in a distributed system as boxes with well-defined interfaces, connections between ports on component interfaces, and specifies component properties that can be used in analytical reasoning about the model. Several architectural modeling languages for embedded systems have emerged in recent years, including the SAE AADL <sup>0</sup>, SysML <sup>0</sup>, UML MARTE <sup>0</sup>.

In system design, an architectural specification serves several important purposes. First, it breaks down a system model into manageable components to establish clear interfaces between components. In this way, complexity becomes manageable by hiding details that are not relevant at a given level of abstraction. Clear, formally defined, component interfaces allow us to avoid integration problems at the implementation phase.

<sup>&</sup>lt;sup>0</sup>Abstract Refinement Types. N. Vazou, P. Rondon, and R. Jhala. European Symposium on Programming. Springer, 2013.

<sup>&</sup>lt;sup>0</sup>LTL types FRP. A. Jeffrey. Programming Languages meets Program Verification.

<sup>&</sup>lt;sup>0</sup>A contract-based formalism for the specification of heterogeneous systems. L. Benvenistu, et al. FDL, 2008

<sup>&</sup>lt;sup>0</sup>Architecture Analysis and Design Language, AS-5506. SAE, 2004. http://standards.sae.org/as5506b

<sup>&</sup>lt;sup>0</sup>System modeling Language. OMG, 2007. http://www.omg.org/spec/SysML

<sup>&</sup>lt;sup>0</sup>UML Profile for MARTE. OMG, 2009. http://www.omg.org/spec/MARTE

Connections between components, which specify how components affect each other, help propagate the effects of a change in one component to the linked components.

Most importantly, an architectural model is a repository to share knowledge about the system being designed. This knowledge can be represented as requirements, design artifacts, component implementations, held together by a structural backbone. Such a repository enables automatic generation of analytical models for different aspects of the system, such as timing, reliability, security, performance, energy, etc. Since all the models are generated from the same source, the consistency of assumptions w.r.t. guarantees, of abstractions w.r.t. refinements, used for different analyses becomes easier, and can be properly ensured in a design methodology based on formal verification and synthesis methods.

Related works in this aim, and closer in spirit to our approach (to focus on modeling time) are domainspecific languages such as Prelude<sup>0</sup> to model the real-time characteristics of embedded software architectures. Conversely, standard architecture description languages could be based on algebraic modeling tools, such as interface theories with the ECDAR tool<sup>0</sup>.

In project TEA, it takes form by the normalization of the AADL standard's formal semantics and the proposal of a time specification annex in the form of related standards, such as CCSL, to model concurrency time and physical properties, and PSL, to model timed traces.

#### 3.4. Scheduling Theory

Based on sound formalization of time and CPS architectures, real-time scheduling theory provides tools for predicting the timing behavior of a CPS which consists of many interacting software and hardware components. Expressing parallelism among software components is a crucial aspect of the design process of a CPS. It allows for efficient partition and exploitation of available resources.

The literature about real-time scheduling <sup>0</sup> provides very mature schedulability tests regarding many scheduling strategies, preemptive or non-preemptive scheduling, uniprocessor or multiprocessor scheduling, etc. Scheduling of data-flow graphs has also been extensively studied in the past decades.

A milestone in this prospect is the development of abstract affine scheduling techniques <sup>0</sup>. It consists, first, of approximating task communication patterns (e.g. between Safety-Critical Java threads) using cyclo-static data-flow graphs and affine functions. Then, it uses state of the art ILP techniques to find optimal schedules and to concretize them as real-time schedules in the program implementations <sup>00</sup>.

Abstract scheduling, or the use of abstraction and refinement techniques in scheduling borrowed to the theory of abstract interpretation <sup>0</sup> is a promising development toward tooled methodologies to orchestrate thousands of heterogeneous hardware/software blocks on modern CPS architectures (just consider modern cars or aircrafts). It is an issue that simply defies the state of the art and known bounds of complexity theory in the field, and consequently requires a particular address.

To develop the underlying theory of this promising research topic, we first need to deepen the theoretical foundation to establish links between scheduling analysis and abstract interpretation. A theory of time systems would offer the ideal framework to pursue this development. It amounts to representing scheduling constraints, inferred from programs, as types or contract properties. It allows to formalize the target time model of the scheduler (the architecture, its middle-ware, its real-time system) and defines the basic concepts to verify assumptions made in one with promises offered by the other: contract verification or, in this case, synthesis.

<sup>&</sup>lt;sup>0</sup>The Prelude language. LIFL and ONERA, 2012. http://www.lifl.fr/~forget/prelude.html

<sup>&</sup>lt;sup>0</sup>PyECDAR, timed games for timed specifications. Inria, 2013. https://project.inria.fr/pyecdar

<sup>&</sup>lt;sup>0</sup>A survey of hard real-time scheduling for multiprocessor systems. R. I. Davis and A. Burns. ACM Computing Survey 43(4), 2011.

<sup>&</sup>lt;sup>0</sup>Buffer minimization in EDF scheduling of data-flow graphs. A. Bouakaz and J.-P. Talpin. LCTES, ACM, 2013.

<sup>&</sup>lt;sup>0</sup>ADFG for the synthesis of hard real-time applications. A. Bouakaz, J.-P. Talpin, J. Vitek. ACSD, IEEE, June 2012.

<sup>&</sup>lt;sup>0</sup>Design of SCJ Level 1 Applications Using Affine Abstract Clocks. A. Bouakaz and J.-P. Talpin. SCOPES, ACM, 2013.

<sup>&</sup>lt;sup>0</sup>La vérification de programmes par interprétation abstraite. P. Cousot. Séminaire au Collège de France, 2008.

#### **3.5. Virtual Prototyping**

Virtual Prototyping is the technology of developing realistic simulators from models of a system under design; that is, an emulated device that captures most, if not all, of the required properties of the real system, based on its specifications. A virtual prototype should be run and tested like the real device. Ideally, the real application software would be run on the virtual prototyping platform and produce the same results as the real device with the same sequence of outputs and reported performance measurements. This may be true to some extent only. Some trade-offs have often to be made between the accuracy of the virtual prototype, and time to develop accurate models.

In order to speed-up simulation time, the virtual prototype must trade-off with something. Depending upon the application designer's goals, one may be interested in trading some loss of accuracy in exchange for simulation speed, which leads to constructing simulation models that focus on some design aspects and provide abstraction of others. A simulation model can provide an abstraction of the simulated hardware in three directions:

- *Computation abstraction*. A hardware component computes a high level function by carrying out a series of small steps executed by composing logical gates. In a virtual prototyping environment, it is often possible to compute the high level function directly by using the available computing resources on the simulation host machine, thus abstracting the hardware function.
- *Communication abstraction*. Hardware components communicate together using some wiring, and some protocol to transmit the data. Simulation of the communication and the particular protocol may be irrelevant for the purpose of virtual prototyping: communication can be abstracted into higher level data transmission functions.
- *Timing Abstraction.* In a cycle accurate simulator, there are multiple simulation tasks, and each task makes some progress on each clock cycle, but this slows down the simulation. In a virtual prototyping experiment, one may not need such precise timing information: coarser time abstractions can be defined allowing for faster simulation.

The cornerstone of a virtual prototyping platform is the component that simulates the processor(s) of the platform, and its associated peripherals. Such simulation can be *static* or *dynamic*.

A solution usually adopted to handle time in virtual prototyping is to manage hierarchical time scales, use component abstractions where possible to gain performance, use refinement to gain accuracy where needed. Localized time abstraction may not only yield faster simulation, but facilitate also verification and synthesis (e.g. synchronous abstractions of physically distributed systems). Such an approach requires computations and communications to be harmoniously discretized and abstracted from originally heterogeneous viewpoints onto a structuring, articulating, pivot model, for concerted reasoning about time and scheduling of events in a way that ensures global system specification correctness.

In the short term these component models could be based on libraries of predefined models of different levels of abstractions. Such abstractions are common in large programming workbench for hardware modeling, such as SystemC, but less so, because of the engineering required, for virtual prototyping platforms.

The approach of team TEA provides an additional ingredient in the form of rich component interfaces. It therefore dictates to further investigate the combined use of conventional virtual prototyping libraries, defined as executable abstractions of real hardware, with executable component simulators synthesised from rich interface specifications (using, e.g., conventional compiling techniques used for synchronous programs).

## 4. Application Domains

#### 4.1. Automotive and Avionics

From our continuous collaboration with major academic and industrial partners through projects TOPCASED, OPENEMBEDD, SPACIFY, CESAR, OPEES, P and CORAIL, our experience has primarily focused on the

aerospace domain. The topics of time and architecture of team TEA extend to both avionics and automotive. Yet, the research focus on time in team TEA is central in any aspect of, cyber-physical, embedded system design in factory automation, automotive, music synthesis, signal processing, software radio, circuit and system on a chip design; many application domains which, should more collaborators join the team, would definitely be worth investigating.

Multi-scale, multi-aspect time modeling, analysis and software synthesis will greatly contribute to architecture modeling in these domains, with applications to optimized (distributed, parallel, multi-core) code generation for avionics (project Corail with Thales avionics, section 8) as well as modeling standards, real-time simulation and virtual integration in automotive (project with Toyota ITC, section 8).

Together with the importance of open-source software, one of these projects, the FUI Project P (section 8), demonstrated that a centralized model for system design could not just be a domain-specific programming language, such as discrete Simulink data-flows or a synchronous language. Synchronous languages implement a fixed model of time using logical clocks that are abstraction of time as sensed by software. They correspond to a fixed viewpoint in system design, and in a fixed hardware location in the system, which is not adequate to our purpose and must be extended.

In project P, we first tried to define a centralized model for importing discrete-continuous models onto a simplified implementation of SIMULINK: P models. Certified code generators would then be developed from that format. Because this does not encompass all aspects being translated to P, the P meta-model is now being extended to architecture description concepts (of the AADL) in order to become better suited for the purpose of system design. Another example is the development of System modeler on top of SCADE, which uses the more model-engineering flavored formalism SysML to try to unambiguously represent architectures around SCADE modules.

An abstract specification formalism, capable of representing time, timing relations, with which heterogeneous models can be abstracted, from which programs can be synthesized, naturally appears better suited for the purpose of virtual prototyping. RT-Builder, based on Signal like Polychrony and developed by TNI, was industrially proven and deployed for that purpose at Peugeot. It served to develop the virtual platform simulating all on-board electronics of PSA cars. This 'hardware in the loop" simulator was used to test equipments supplied by other manufacturers with respect to virtual cars. In the advent of the related automotive standard, RT-Builder then became AUTOSAR-Builder.

#### 4.2. Factory Automation

In collaboration with Mitsubishi R&D, we explore another application domain where time and domain heterogeneity are prime concerns: factory automation. In factory automation alone, a system is conventionally built from generic computing modules: PLCs (Programmable Logic Controllers), connected to the environment with actuators and detectors, and linked to a distributed network. Each individual, physically distributed, PLC module must be timely programmed to perform individually coherent actions and fulfill the global physical, chemical, safety, power efficiency, performance and latency requirements of the whole production chain. Factory chains are subject to global and heterogeneous (physical, electronic, functional) requirements whose enforcement must be orchestrated for all individual components.

Model-based analysis in factory automation emerges from different scientific domains and focus on different CPS abstractions that interact in subtle ways: logic of PLC programs, real-time electromechanical processing, physical and chemical environments. This yields domain communication problems that render individual domain analysis useless. For instance, if one domain analysis (e.g. software) modifies a system model in a way that violates assumptions made by another domain (e.g. chemistry) then the detection of its violation may well be impossible to explain to either of the software and chemistry experts. As a consequence, cross-domain analysis issues are discovered very late during system integration and lead to costly fixes. This is particularly prevalent in multi-tier industries, such as avionic, automotive, factories, where systems are prominently integrated from independently-developed parts.

## 5. Highlights of the Year

#### 5.1. Highlights of the Year

Inria created a new International Chair and appointed American computer engineer Rajesh Gupta to the part-time position. Gupta is a professor and former chair of the Computer Science and Engineering (CSE) department in the Jacobs School of Engineering at the University of California San Diego. Rajesh Gupta will hold the International Chair for a period of five years. Starting this summer, he will engage with researchers in Inria's research center in Rennes. The position enables him to spend as much as a year spread out over the five years of his appointment.

## 6. New Software and Platforms

#### 6.1. ADFG

Affine data-flow graphs schedule synthesizer

KEYWORDS: Code generation - Scheduling - Static program analysis

FUNCTIONAL DESCRIPTION: ADFG is a synthesis tool of real-time system scheduling parameters: ADFG computes task periods and buffer sizes of systems resulting in a trade-off between throughput maximization and buffer size minimization. ADFG synthesizes systems modeled by ultimately cyclo-static dataflow (UCSDF) graphs, an extension of the standard CSDF model.

Knowing the WCET (Worst Case Execute Time) of the actors and their exchanges on the channels, ADFG tries to synthezise the scheduler of the application. ADFG offers several scheduling policies and can detect unschedulable systems. It ensures that the real scheduling does not cause overflows or underflows and tries to maximize the throughput (the processors utilization) while minimizing the storage space needed between the actors (i.e. the buffer sizes).

Abstract affine scheduling is first applied on the dataflow graph, that consists only of periodic actors, to compute timeless scheduling constraints (e.g. relation between the speeds of two actors) and buffering parameters. Then, symbolic schedulability policies analysis (i.e., synthesis of timing and scheduling parameters of actors) is applied to produce the scheduller for the actors.

ADFG, initially defined to synthesize real-time schedulers for SCJ/L1 applications, may be used for scheduling analysis of AADL programs.

- Authors: Thierry Gautier, Jean-Pierre Talpin, Adnan Bouakaz, Alexandre Honorat and Loïc Besnard
- Contact: Loïc Besnard

#### **6.2. POLYCHRONY**

KEYWORDS: Code generation - AADL - Proof - Optimization - Multi-clock - GALS - Architecture - Cosimulation - Real time - Synchronous Language

FUNCTIONAL DESCRIPTION: Polychrony is an Open Source development environment for critical/embedded systems. It is based on Signal, a real-time polychronous data-flow language. It provides a unified modeldriven environment to perform design exploration by using top-down and bottom-up design methodologies formally supported by design model transformations from specification to implementation and from synchrony to asynchrony. It can be included in heterogeneous design systems with various input formalisms and output languages. The Polychrony tool-set provides a formal framework to: validate a design at different levels, by the way of formal verification and/or simulation, refine descriptions in a top-down approach, abstract properties needed for black-box composition, compose heterogeneous components (bottom-up with COTS), generate executable code for various architectures. The Polychrony tool-set contains three main components and an experimental interface to GNU Compiler Collection (GCC): \* The Signal toolbox, a batch compiler for the Signal language, and a structured API that provides a set of program transformations. Itcan be installed without other components and is distributed under GPL V2 license.

\* The Signal GUI, a Graphical User Interface to the Signal toolbox (editor + interactive access to compiling functionalities). It can be used either as a specific tool or as a graphical view under Eclipse. It has been transformed and restructured, in order to get a more up-to-date interface allowing multi-window manipulation of programs. It is distributed under GPL V2 license.

\* The POP Eclipse platform, a front-end to the Signal toolbox in the Eclipse environment. It is distributed under EPL license.

- Participants: Loïc Besnard, Paul Le Guernic and Thierry Gautier
- Partners: CNRS Inria
- Contact: Loïc Besnard
- URL: https://www.polarsys.org/projects/polarsys.pop

#### 6.3. Polychrony AADL2SIGNAL

KEYWORDS: Real-time application - Polychrone - Synchronous model - Polarsys - Polychrony - Signal - AADL - Eclipse - Meta model

FUNCTIONAL DESCRIPTION: This polychronous MoC has been used previously as semantic model for systems described in the core AADL standard. The core AADL is extended with annexes, such as the Behavior Annex, which allows to specify more precisely architectural behaviors. The translation from AADL specifications into the polychronous model should take into account these behavior specifications, which are based on description of automata.

For that purpose, the AADL state transition systems are translated as Signal automata (a slight extension of the Signal language has been defined to support the model of polychronous automata).

Once the AADL model of a system transformed into a Signal program, one can analyze the program using the Polychrony framework in order to check if timing, scheduling and logical requirements over the whole system are met.

We have implemented the translation and experimented it using a concrete case study, which is the AADL modeling of an Adaptive Cruise Control (ACC) system, a highly safety-critical system embedded in recent cars.

- Participants: Huafeng Yu, Loïc Besnard, Paul Le Guernic, Thierry Gautier and Yue Ma
- Partner: CNRS
- Contact: Loïc Besnard
- URL: http://www.inria.fr/equipes/tea

#### 6.4. POP

#### Polychrony on Polarsys

KEYWORDS: Synchronous model - Model-driven engineering

FUNCTIONAL DESCRIPTION: The Eclipse project POP is a model-driven engineering front-end to our opensource toolset Polychrony. a major achievement of the ESPRESSO (and now TEA) project-team. The Eclipse project POP is a model-driven engineering front-end to our open-source toolset Polychrony. It was finalised in the frame of project OPEES, as a case study: by passing the POLARSYS qualification kit as a computer aided simulation and verification tool. This qualification was implemented by CS Toulouse in conformance with relevant generic (platform independent) qualification documents. Polychrony is now distributed by the Eclipse project POP on the platform of the POLARSYS industrial working group. Team TEA aims at continuing its dissemination to academic partners, as to its principles and features, and industrial partners, as to the services it can offer. Project POP is composed of the Polychrony tool set, under GPL license, and its Eclipse framework, under EPL license. SSME (Syntactic Signal-Meta under Eclipse), is the meta-model of the Signal language implemented with Eclipse/Ecore. It describes all syntactic elements specified in Signal Reference Manual <sup>0</sup>: all Signal operators (e.g. arithmetic, clock synchronization), model (e.g. process frame, module), and construction (e.g. iteration, type declaration). The meta-model primarily aims at making the language and services of the Polychrony environment available to inter-operation and composition with other components (e.g. AADL, Simulink, GeneAuto, P) within an Eclipse-based development tool-chain. Polychrony now comprises the capability to directly import and export Ecore models instead of textual Signal programs, in order to facilitate interaction between components within such a tool-chain. The download site for project POP has opened in 2015 at https://www.polarsys.org/projects/polarsys.pop. It should be noted that the Eclipse Foundation does not host code under GPL license. So, the Signal toolbox useful to compile Signal code from Eclipse is hosted on our web server.

- Participants: Jean-Pierre Talpin, Loïc Besnard, Paul Le Guernic and Thierry Gautier
- Contact: Loïc Besnard
- URL: https://www.polarsys.org/projects/polarsys.pop

#### 6.5. Sigali

FUNCTIONAL DESCRIPTION: Sigali is a model-checking tool that operates on ILTS (Implicit Labeled Transition Systems, an equational representation of an automaton), an intermediate model for discrete event systems. It offers functionalities for verification of reactive systems and discrete controller synthesis. The techniques used consist in manipulating the system of equations instead of the set of solutions, which avoids the enumeration of the state space. Each set of states is uniquely characterized by a predicate and the operations on sets can be equivalently performed on the associated predicates. Therefore, a wide spectrum of properties, such as liveness, invariance, reachability and attractivity, can be checked. Algorithms for the computation of predicates on states are also available. Sigali is connected with the Polychrony environment (Tea project-team) as well as the Matou environment (VERIMAG), thus allowing the modeling of reactive systems by means of Signal Specification or Mode Automata and the visualization of the synthesized controller by an interactive simulation of the controlled system.

• Contact: Hervé Marchand

## 7. New Results

#### 7.1. ADFG: Affine data-flow graphs scheduler synthesis

Participants: Loïc Besnard, Thierry Gautier, Alexandre Honorat, Jean-Pierre Talpin, Hai Nam Tran.

We consider with ADFG (Affine DataFlow Graph) the synthesis of periodic scheduling parameters for realtime systems modeled as ultimately cyclo-static dataflow (UCSDF) graphs [14]. This synthesis aims for a trade-off between throughput maximization and total buffer size minimization. The synthesizer inputs are: a UCSDF graph which describes tasks by their Worst Case Execution Time (WCET), and directed buffers connecting tasks by their data production and consumption rates; the number of processors in the target system and the real-time scheduling synthesis algorithm to be used. The outputs are the synthesized scheduling parameters: the tasks periods, offsets, processor bindings and priorities, and the buffers initial marking and maximum sizes.

<sup>0</sup> 

*SIGNAL V4-Inria version: Reference Manual.* Besnard, L., Gautier, T. and Le Guernic, P. http://www.irisa.fr/espresso/Polychrony, 2010

ADFG was originally the implementation of Adnan Bouakaz's work <sup>0</sup>. However the tool had not been packaged yet to be easily installed and used. Moreover, code refactoring led to improve the theory, and to add new features. Firstly, more accurate bounds and Integer Linear Programming (ILP) formulations have been used. Besides, dataflow graphs do not need to be weakly connected for EDF policy on multiprocessor systems. The new implementation also avoids to use a fixed parameter for some multiprocessor partitioning algorithms, now an optional strategy enables to compute it. Finally implementation has been adapted to standard technologies to be more easily installed and used. As the synthesizer evolved a lot, new evaluations have been made. Moreover, many scheduled examples have been simulated with Cheddar <sup>0</sup>, which provides pertinent metrics to analyze the scheduling efficiency.

ADFG is being extended to investigate and solve the scheduling problem of dataflow programs on many-core architectures. These architectures have distinctive traits requiring significant changes to classical multiprocessor scheduling theory. There is a high number of contention points introduced by novel memory architectures and new interconnect types such as Network-on-Chip. Two solutions are proposed and implemented in ADFG: contention-aware and contention-free scheduling synthesis. We either take into account the contention and synthesize a contention-aware schedule or find a schedule that results in no contention.



Figure 1. ADFG under Eclipse

#### 7.2. Formal Semantics of Behavior Specifications in the Architecture Analysis and Design Language Standard

Participants: Loïc Besnard, Thierry Gautier, Jean-Pierre Talpin.

The Architecture Analysis and Design Language (AADL) is a standard proposed by SAE to express architecture specifications and share knowledge between the different stakeholders about the system being designed. To support unambiguous reasoning, formal verification, high-fidelity simulation of architecture specifications in a model-based AADL design workflow, we have defined formal semantics for the behavior specification of the AADL. These semantics rely on the structure of automata present in the standard already, yet provide

<sup>&</sup>lt;sup>0</sup>Real-Time Scheduling of Dataflow Graphs. A. Bouakaz. PhD Thesis, University of Rennes 1, 2013.

<sup>&</sup>lt;sup>0</sup>The Cheddar project: a GPL real-time scheduling analyzer.http://beru.univ-brest.fr/~singhoff/cheddar/

tagged, trace semantics framework to establish formal relations between (synchronous, timed, asynchronous) usages or interpretations of behavior [17]. We define the model of computation and communication of a behavior specification by the synchronous, timed or asynchronous traces of automata with variables. These constrained automata are derived from *polychronous automata* defined within the polychronous model of computation and communication [11].

States of a behavior annex transition system can be either observable from the outside (*initial, final* or *complete* states), that is states in which the execution of the component is paused or stopped and its outputs are available; or non observable execution states, that is internal states. We thus define two kinds of steps in the transition system: *small steps*, that is non-observable steps from or to an internal state; and *big steps*, that is observable steps from a *complete* state to another, through a number of small steps). The semantics of the AADL considers the observable states of the automaton. The set of states  $S_A$  of automaton A (used to interpret the behavior annex) thus only contains states corresponding to these observable states and the set of transitions from or to an execution state). The action language of the behavior annex defines actions performed during transitions. Actions associated with transitions are action blocks that are built from basic actions and a minimal set of control structures (sequences, sets, conditionals and loops). Typically, a behavior action set is represented by concatenating the transition systems of its elements; a behavior action set is represented by composing the transition systems of its elements.

The polychronous model of computation had been used previously as semantic model for systems described in the core AADL standard. This translation of AADL specifications into the polychronous model now takes into account the behavior specifications. The import of AADL behavior annexes (AADL-BA) to the polychronous model relies on polychronous automata and on small steps/big steps semantics. Small steps may be viewed as an implicit oversampling of the big steps. To express such implicit upsampling, a model of *Signal-thread* has been introduced in Polychrony (refer to Section "New trends and developments in Polychrony"). In that context, the translation of a behavior annex associated with an AADL thread consists mainly in the production of the corresponding Signal automaton, which is declared as a Signal-thread, and the definition of the environment required for this Signal-thread. In particular, the signal *complete-thread* is defined so that it will occur when the next state of the automaton is a *complete* state (the control will return to the scheduler): in other words, it specifies the end of a sequence of small steps.

A specific difficulty in the translation of AADL-BA is the translation of the action language, which is related to the general problem of the translation of a sequential language to a dataflow one. First, in AADL-BA actions, a given variable may be assigned several times in a sequence (for example, x = a + b; x = x + a). Thus an AADL-BA action has to be transformed into a SSA (static single assignment) form ( $x_0 = a + b$ ;  $x = x_0 + a$  in the previous example). Another possible problem is the translation of AADL-BA loop structures (for, while, do until). In our case, this is solved, again, by considering them as Signal-threads: the *dispatch-thread* event is defined by the upperbound of the clocks of the inputs of the loop and the *complete-thread* event defines the termination of the loop.

#### 7.3. New trends and developments in Polychrony

Participants: Loïc Besnard, Thierry Gautier.

The synchronous modeling paradigm provides strong correctness guarantees for embedded system design while requiring minimal environmental assumptions. In most related frameworks, global execution correctness is achieved by ensuring the insensitivity of (logical) time in the program from (real) time in the environment. This property, called endochrony, can be statically checked, making it fast to ensure design correctness. Unfortunately, it is not preserved by composition, which makes it difficult to exploit with component-based design concepts in mind.

It has been shown that compositionality can be achieved by weakening the objective of endochrony: a weakly endochronous system is a deterministic system that can perform independent computations and communications in any order as long as this does not alter its global state. Moreover, the non-blocking

composition of weakly endochronous processes is isochronous, which means that the synchronous and asynchronous compositions of weakly endochronous processes accept the same behaviors. Unfortunately, testing weak endochrony needs state-space exploration, which is very costly in the general case. Then, a particular case of weak endochrony, called polyendochrony, was defined, which allows static checking thanks to the existing clock calculus. The clock hierarchy of a polyendochronous system may have several trees, with synchronization relations between clocks placed in different trees, but the clock expressions of the clock system must be such that there is no clock expression (especially, no root clock expression) defined by symmetric difference: root clocks cannot refer to absence. In other words, the clock system must be in disjunctive form [9].

We have now implemented code generation for polyendochronous systems in Polychrony. This generation reuses techniques of distributed code generation, with rendez-vous management for synchronization constraints on clocks which are not placed in the same tree of clocks. For such a synchronization constraint  $c_1 = c_2$ , nodes *send* and *receive* are added in the graph, associated with clocks  $c_1$  and  $c_2$ : for  $c_1$ , *send*( $c_1$ ) is followed by *receive*( $c_2$ ), followed itself by all the other nodes associated with clock  $c_1$ ; and symmetrically for  $c_2$ . Then the subgraphs corresponding respectively to the trees where  $c_1$  and  $c_2$  are placed are separated, as if they were distributed on different processors. In this way, nodes *send* and *receive* become respectively outputs and inputs (both for  $c_1$  and  $c_2$ ) of the subgraphs. Finally, a communication library (MPI) is used for simulation. The following restriction is considered in the current implementation: the roots of the trees of  $c_1$ and  $c_2$  must be free variables.

We have also considered another extension related to clocks, again for making code generation possible for more programs than it was the case before. A characteristic of the Signal language is that it allows to specify programs which have internal accelerations with respect to their inputs and outputs. However, the constraint that implemented programs, for which code was generated, should be endochronous, restricted more or less these programs to have one single such acceleration (or clock upsampling). To abstract from this restriction, we have defined a model of so-called *Signal-thread*, that helps to confine such accelerations, and thus to generate code for programs with multiple clock upsampling. A Signal-thread is a Signal process with internal implicit upsampling; it has a *dispatch-thread* input event and a *complete-thread* output event; its outputs are delayed compared with its inputs. As the Signal-thread represents an upsampling, the *step* (see [1]) of the corresponding generated code is a loop. Such Signal-threads may be considered as a pragmatic way to implement *clock domains*.

#### 7.4. Modular verification of cyber-physical systems using contract theory

Participants: Jean-Pierre Talpin, Benoit Boyer, David Mentre, Simon Lunel.

The primary goal of our project, in collaboration with Mitsubishi Electronics Research Centre Europe (MERCE), is to ensure correctness-by-design in realistic cyber-physical systems, i.e., systems that mix software and hardware in a physical environment, e.g., Mitsubishi factory automation lines or water-plant factory. To achieve that, we develop a verification methodology based on decomposition into components enhanced with contract reasoning.

The work of A. Platzer on Differential Dynamic Logic  $(d\mathcal{L})$  holds our attention <sup>0</sup>. This a formalism built on the Dynamic Logic of V. Pratt augmented with the possibility of expressing Ordinary Differential Equations (ODEs), which are the usual way to model physical behaviors in physics. Combined with the ability of Dynamic Logic to specify and verify hybrid programs,  $d\mathcal{L}$  is particularly fit model cyber-physical systems. The proof system associated with the logic is implemented into the theorem prover KeYmaera X. Aimed toward automatisation, it is a promising tool to spread formal methods into industry.

We have defined a syntactic parallel composition operator in  $d\mathcal{L}$  which enjoys associativity and commutativity [15]. Commutativity allows to compose component in every possible order. Associativity is mandatory to modularly design a system; it allow to upgrade a system by adding new components. We have then characterized the conditions under which we can derive automatically a proof of the contract of our composition of two

<sup>&</sup>lt;sup>0</sup>Differential Dynamic Logic for Hybrid Systems, André Platzer, http://symbolaris.com/logic/dL.html

components, given the proof of the contract for each components. Theses theoretical results have been exemplified with an example of a cruise-controller entirely proved within the interactive theorem prover KeYmaera X.

The study of the cruise-controller example and of a water-tank system highlights some limitations of our approach. We can not handle retro-action and we have to compose in parallel components which have to be sequenced, e.g. a sensor and a computer. We have overcomed theses limitations by introducing a sequential composition operator which enjoys associativity and distributivity over the parallel composition operator. We believe it is a first step toward a composition algebra in  $d\mathcal{L}$ . This operator also satisfy the property that we can automatically derive a proof of the contract of our composition of two components, given the proof of the contract for each components, but under some relaxed conditions. We believe it is the first step toward a composition algebra.

Thanks to these results, a wide variety of systems are now possible to modularly design in  $d\mathcal{L}$ . To validate our approach, we are currently working on the implementation of our parallel composition operator as a tactic in KeYmaera X.

To challenge our ideas, we are working in the proof of a realistic cyber-physical system, a power-train system used in automotive. We plan to use it as a basis to test abstraction mechanisms to ultimately allow mix between top-down and bottom-up design.

#### 7.5. Parametric verification of time synchronization protocols

Participants: Ocan Sankur, Jean-Pierre Talpin.

In the context of the associate-team COMPOSITE, we addressed the verification one of the apparently simplest services in any loosely-coupled distributed system : the time service. In many instances of such systems, traffic and power grids, banking and transaction networks, the accuracy and reliability of this service are critical.

In the instance of sensor networks, it is of particular interest to verify the robustness of such protocols to variations caused by the environment. Lake of power, varying temperatures, imperfect hardware, are sources of local drifts and jitters in time measurement that require self-calibration and fault-tolerance to reach distributed consensus. FTSP, the flooding time synchronization protocol, provides fault-tolerance and enables time synchronization.

In [16], we introduce an environment abstraction technique and an incremental model checking technique to prove that FTSP eventually elects a leader for any network topology and configuration (anonymized identifiers), up to a diameter N = 7 (with synchronous communications) and N = 5 (desynchronized communications), resulting in significant improvements over previous results.

#### 7.6. Modular analysis and verification of system libraries

Participants: Jean-Joseph Marty, Jean-Pierre Talpin.

We are starting to develop a new perspective on the active topic of information flow control (IFC). We plan to adapt current investigations to tagged multi-core architecture, including software (virtual machines) and hardware (the Risc V processor) experiments and applications. All this work is based on the previous experience about verified Unikernel programming on low resources processors such as the Arduino (Marty's Master internship). We will define formally relations between processes and blocks of code inside a concurrent environment. This line of work will be investigated for both embedded IoT applications and cloud computing. By working with IFC at processor level and system level, we will enforce strong security foundation and focus on constraint solving analysed software.

## 8. Bilateral Contracts and Grants with Industry

#### 8.1. Bilateral Grants with Industry

#### 8.1.1. Mitsubishi Electric R&D Europe (2015-2018)

Title: Analysis and verification for correct by construction orchestration in automated factories Inria principal investigator: Jean-Pierre Talpin, Simon Lunel International Partner: Mitsubishi Electric R&D Europe Duration: 2015 - 2018

Abstract: The primary goal of our project is to ensure correctness-by-design in cyber-physical systems, i.e., systems that mix software and hardware in a physical environment, e.g., Mitsubishi factory automation lines. We develop a component-based approach in Differential Dynamic Logic allowing to reason about a wide variety of heterogeneous cyber-physical systems. Our work provides tools and methodology to design and prove a system modularly.

## 9. Partnerships and Cooperations

#### 9.1. National Initiatives

#### 9.1.1. ANR

Program: ANR

Project acronym: Feever

Project title: Faust Environment Everyware

Duration: 2014-2016

Coordinator: Pierre Jouvelot, Mines ParisTech

Other partners: Grame, Inria Rennes, CIEREC

URL: http://www.feever.fr

Abstract: The aim of project FEEVER is to ready the Faust music synthesis language for the Web. In this context, we collaborate with Mines ParisTech to define a type system suitable to model music signals timed at multiple rates and to formally support playing music synthesized from different physical locations.

#### 9.1.2. PAI

Program: PAI/CORAC

Project acronym: CORAIL

Project title: Composants pour l'Avionique Modulaire Étendue

Duration: July 2013 - May 2017

Coordinator: Thales Avionics

Other partners: Airbus, Dassault Aviation, Eurocopter, Sagem ...

Abstract: The CORAIL project aims at defining components for Extended Modular Avionics. The contribution of project-team TEA is to define a specification method and to provide a generator of multi-task applications.

#### 9.2. International Initiatives

#### 9.2.1. Inria International Labs

9.2.1.1. SACCADES

Title: Saccades International Partner: LIAMA East China Normal University Inria project-teams Aoste and Tea

Duration: 2003 - now

The SACCADES project is a LIAMA project hosted by East China Normal University and jointly led by Vania Joloboff (Inria) and Min Zhang (ECNU). The SACCADES project aims at improving the development of reliable cyber physical systems and more generally of distributed systems combining asynchronous with synchronous aspects, with different but complementary angles:

- develop the theoretical support for Models of Computations and Communications (MoCCs) that are the fundamentals basis of the tools.
- develop software tools (a) to enable the development and verification of executable models of the application software, which may be local or distributed and (b) to define and optimize the mapping of software components over the available resources.
- develop virtual prototyping technology enabling the validation of the application software on the target hardware platform.

The ambition of SACCADES project is to develop

- Theoretical Support for Cyber Physical Systems
- Software Tools for design and validation of CPS
- Virtual Prototyping of CPS

#### 9.2.2. Inria Associate Teams

9.2.2.1. Composite

Title: Compositional System Integration

International Partner (Institution - Laboratory - Researcher):

• University of California, San Diego (United States) - Microelectronic Embedded Systems Laboratory - Rajesh Gupta

Start year: 2017

See also: http://www.irisa.fr/prive/talpin/composite

Most applications that run somewhere on the internet are not optimized to do so. They execute on general purpose operating systems or on containers (virtual machines) that are built with the most conservative assumptions about their environment. While an application is specific, a large part of the system it runs on is unused, which is both a cost (to store and execute) and a security risk (many entry points).

A unikernel, on the contrary, is a system program object that only contains the necessary the operating system services it needs for execution. A unikernel is build from the composition of a program, developed using high-level programming language, with modules of a library operating system (libOS), to execute directly on an hypervisor. A unikernel can boot in milliseconds to serve a request and shut down, demanding minimal energy and resources, offering stealthiest exposure time and surface to attacks, making them the ideal platforms to deploy on sensor networks, networks of embedded devices, smart grids and clouds.

The goal of COMPOSITE is to develop the mathematical foundations for sound and efficient composition in system programming: analysis, verification and optimization technique for modular and compositional hardware-system-software integration of unikernels. We intend to further this development with the prospect of an end-to-end co-design methodology to synthesize lean and stealth networked embedded devices.

#### 9.2.3. Inria International Partners

#### 9.2.3.1. Convex

Title: Compositional Verification of Cyber-Physical Systems International Partner:

- Chinese Academy of Science, Institute of Software
- Beihang University
- Nanhang University
- Nankai University

Duration: 2017 - now

Formal modeling and verification methods have successfully improved software safety and security in vast application domains in transportation, production and energy. However, formal methods are labor-intensive and require highly trained software developers. Challenges facing formal methods stem from rapid evolution of hardware platforms, the increasing amount and cost of software infrastructures, and from the interaction between software, hardware and physics in networked cyber-physical systems.

Automation and expressivity of formal verification tools must be improved not only to scale functional verification to very large software stacks, but also verify non-functional properties from models of hardware (time, energy) and physics (domain). Abstraction, compositionality and refinement are essential properties to provide the necessary scalability to tackle the complexity of system design with methods able to scale heterogeneous, concurrent, networked, timed, discrete and continuous models of cyber-physical systems.

Project Convex wants to define a CPS architecture design methodology that takes advantage of existing time and concurrency modeling standards (MARTE, AADL, Ptolemy, Matlab), yet focuses on interfacing heterogeneous and exogenous models using simple, mathematically-defined structures, to achieve the single goal of correctly integrating CPS components.

#### 9.3. International Research Visitors

#### 9.3.1. Visits of International Scientists

Rajesh Gupta visited project-team TEA in August and gave two 68NQTR seminars on "Building Computing Machines That Sense, Adapt and Approximate" and on "Compositional Synthesis for High-level Design of System-Chips".

Deian Stefan visited project-team TEA in September and gave a 68NQTR seminar on "Practical multi-core information flow control"

Shuvra Bhattacharyya visited project-team TEA in August and December and gave a 68NQTR seminar on "The DSPCAD Framework for Dataflow-based Design and Implementation of Signal Processing Systems"

#### 9.3.2. Visits to International Teams

Jean-Pierre Talpin visited UC San Diego and UC Berkeley in the context of the associate-project Composite in June.

In the context of the IIP Convex, Jean-Pierre Talpin was invited at Beihang and Nanhang Universities in April, visited Beihang and Nankai Universities in July, and Beihang, Nankai and ECNU in November, to give seminars and a introductory course on model checking.

Jean-Pierre Talpin gave an invited talk on "Parametric model-checking the FTSP protocol" at TU Wien June 30.

Simon Lunel visited CMU and UC San Diego in December to give seminars on "compositional proofs in differential dynamic logic".

## **10.** Dissemination

#### **10.1. Promoting Scientific Activities**

#### 10.1.1. Scientific events organisation

#### 10.1.1.1. General chair, scientific chair

Jean-Pierre Talpin served as General Chair and Finance Chair of the 15th. ACM-IEEE Conference on Methods and Models for System Design in Vienna.

#### 10.1.1.2. Member of the organizing committees

Jean-Pierre Talpin is a member of the steering committee of the ACM-IEEE Conference on Methods and Models for System Design (MEMOCODE).

#### 10.1.2. Scientific Events Selection

#### 10.1.2.1. Member of the Conference Program Committees

Jean-Pierre Talpin served the program committee of:

- ACSD'17, 17th. International Conference on Application of Concurrency to System Design
- LCTES'17, 20th. ACM SIGPLAN-SIGBED Conference on Languages, Compilers, and Tools for Embedded Systems
- SAC'18, 33rd. ACM SIGAPP Symposium on Applied Computing
- SCOPES'17, 20th. International Workshop on Software and Compilers for Embedded Systems

#### 10.1.3. Journal

#### 10.1.3.1. Member of the Editorial Boards

Jean-Pierre Talpin is Associate Editor with the ACM Transactions for Embedded Computing Systems (TECS).

#### 10.1.3.2. Reviews

Thierry Gautier reviewed articles for Information Processing Letters.

#### 10.1.4. Invited Talks

Jean-Pierre Talpin gave a keynote speech, entitled "Compositional methods for CPS design" at the Symposium on Dependable Software Engineering (SETTA'17) in Changsha, October 25.

#### 10.1.5. Invited Talks

Albert Benveniste and Thierry Gautier previewed a seminar at SYNCHRON'17 to be given at the Collège de France in Gérard Berry's 2017-2018 lecture course.

#### 10.1.6. Scientific Expertise

Jean-Pierre Talpin was nominated vice-president of ANR evaluation committee CES-25

#### 10.2. Teaching - Supervision - Juries

#### 10.2.1. Teaching

Jean-Pierre Talpin gave a one week class "Introduction to model-checking" at Nankai University in July and at Beihang University in November.

#### 10.2.2. Supervision

Jean-Pierre Talpin co-supervises the PhD Thesis of Simon Lunel, Liangcong Zhang and Jean-Joseph Marty

#### 10.2.3. Juries

Jean-Pierre Talpin served as rapporteur at the HDR Thesis defense of Jérôme Hugues, entitled "Architecture in the Service of Real-Time Middleware, Contributions to Architecture Description Languages", which took place at INP Toulouse on February 22.

Jean-Pierre Talpin served as rapporteur at the HDR Thesis defense of Maxime Pelcat, entitled "Models, methods and tools for bridging the design productivity gap of embedded signal processing systems", which took place at Institut Poincarré in Clermond-Ferrand on July 10.

## 11. Bibliography

#### Major publications by the team in recent years

- [1] L. BESNARD, T. GAUTIER, P. LE GUERNIC, J.-P. TALPIN. Compilation of Polychronous Data Flow Equations, in "Synthesis of Embedded Software", S. K. SHUKLA, J.-P. TALPIN (editors), Springer, 2010, p. 1-40 [DOI: 10.1007/978-1-4419-6400-7\_1], http://hal.inria.fr/inria-00540493.
- [2] A. BOUAKAZ, J.-P. TALPIN. Design of Safety-Critical Java Level 1 Applications Using Affine Abstract Clocks, in "International Workshop on Software and Compilers for Embedded Systems", St. Goar, Germany, June 2013, p. 58-67 [DOI: 10.1145/2463596.2463600], https://hal.inria.fr/hal-00916487.
- [3] C. BRUNETTE, J.-P. TALPIN, A. GAMATIÉ, T. GAUTIER. A metamodel for the design of polychronous systems, in "The Journal of Logic and Algebraic Programming", 2009, vol. 78, n<sup>o</sup> 4, p. 233 - 259, IFIP WG1.8 Workshop on Applying Concurrency Research in Industry [DOI : 10.1016/J.JLAP.2008.11.005], http:// www.sciencedirect.com/science/article/pii/S1567832608000957.
- [4] A. GAMATIÉ, T. GAUTIER, P. LE GUERNIC, J.-P. TALPIN. Polychronous Design of Embedded Real-Time Applications, in "ACM Transactions on Software Engineering and Methodology (TOSEM)", April 2007, vol. 16, n<sup>o</sup> 2, http://doi.acm.org/10.1145/1217295.1217298.
- [5] A. GAMATIÉ, T. GAUTIER. The Signal Synchronous Multiclock Approach to the Design of Distributed Embedded Systems, in "IEEE Transactions on Parallel and Distributed Systems", 2010, vol. 21, n<sup>o</sup> 5, p. 641-657 [DOI: 10.1109/TPDS.2009.125], http://hal.inria.fr/inria-00522794.
- [6] A. GAMATIÉ, T. GAUTIER, P. LE GUERNIC. Synchronous design of avionic applications based on model refinements, in "Journal of Embedded Computing (IOS Press)", 2006, vol. 2, n<sup>o</sup> 3-4, p. 273-289, http://hal. archives-ouvertes.fr/hal-00541523.
- [7] P. LE GUERNIC, J.-P. TALPIN, J.-C. LE LANN. Polychrony for system design, in "Journal of Circuits, Systems and Computers, Special Issue on Application Specific Hardware Design", June 2003, vol. 12, n<sup>o</sup> 03, http:// hal.inria.fr/docs/00/07/18/71/PDF/RR-4715.pdf.
- [8] D. POTOP-BUTUCARU, Y. SOREL, R. DE SIMONE, J.-P. TALPIN. From Concurrent Multi-clock Programs to Deterministic Asynchronous Implementations, in "Fundamenta Informaticae", January 2011, vol. 108, n<sup>o</sup> 1-2, p. 91–118, http://dl.acm.org/citation.cfm?id=2362088.2362094.
- [9] J.-P. TALPIN, J. OUY, T. GAUTIER, L. BESNARD, P. LE GUERNIC. Compositional design of isochronous systems, in "Science of Computer Programming", February 2012, vol. 77, n<sup>o</sup> 2, p. 113-128 [DOI: 10.1016/J.SCICO.2010.06.006], http://hal.archives-ouvertes.fr/hal-00768341.

[10] H. YU, J. PRASHI, J.-P. TALPIN, S. K. SHUKLA, S. SHIRAISHI. Model-Based Integration for Automotive Control Software, in "Digital Automation Conference", San Francisco, United States, ACM, June 2015, https:// hal.inria.fr/hal-01148905.

#### **Publications of the year**

#### **Articles in International Peer-Reviewed Journal**

- [11] T. GAUTIER, C. GUY, A. HONORAT, P. LE GUERNIC, J.-P. TALPIN, L. BESNARD. Polychronous Automata and their Use for Formal Validation of AADL Models, in "Frontiers of Computer Science", 2017 [DOI: 10.1007/s11704-017-6134-5], https://hal.inria.fr/hal-01411257.
- [12] K. HU, T. ZHANG, L. SHANG, Z. YANG, J.-P. TALPIN. Parallel Code Generation from Synchronous Specifications, in "Journal of Software", November 2017, vol. 28, p. 1-15 [DOI: 10.13328/J.CNKI.JOS.005056], https://hal.inria.fr/hal-01644290.

#### **Invited Conferences**

[13] J.-P. TALPIN. Compositional methods for CPS design (keynote abstract), in "3rd Symposium on Dependable Software Engineering: Theories, Tools and Applications", Changsha, China, Springer, October 2017, https:// hal.inria.fr/hal-01615148.

#### **International Conferences with Proceedings**

- [14] A. HONORAT, H. N. TRAN, L. BESNARD, T. GAUTIER, J.-P. TALPIN, A. BOUAKAZ.ADFG: a scheduling synthesis tool for dataflow graphs in real-time systems, in "International Conference on Real-Time Networks and Systems", Grenoble, France, October 2017, p. 1-10 [DOI : 10.1145/3139258.3139267], https://hal. inria.fr/hal-01615142.
- [15] S. LUNEL, B. BOYER, J.-P. TALPIN. Compositional proofs in differential dynamic logic dL, in "17th International Conference on Application of Concurrency to System Design", Zaragoza, Spain, June 2017, https://hal.inria.fr/hal-01615140.
- [16] O. SANKUR, J.-P. TALPIN. An Abstraction Technique For Parameterized Model Checking of Leader Election Protocols: Application to FTSP, in "23rd International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)", Uppsala, Sweden, Lecture Notes in Computer Science, April 2017, vol. 10206, https://hal.archives-ouvertes.fr/hal-01431472.

#### Scientific Books (or Scientific Book chapters)

- [17] L. BESNARD, T. GAUTIER, P. LE GUERNIC, C. GUY, J.-P. TALPIN, B. LARSON, E. BORDE.Formal Semantics of Behavior Specifications in the Architecture Analysis and Design Language Standard, in "Cyber-Physical System Design from an Architecture Analysis Viewpoint", Cyber-Physical System Design from an Architecture Analysis Viewpoint, Springer, January 2017 [DOI: 10.1007/978-981-10-4436-6\_3], https:// hal.inria.fr/hal-01615143.
- [18] S. NAKAJIMA, J.-P. TALPIN, M. TOYOSHIMA, H. YU. Cyber-Physical System Design from an Architecture Analysis Viewpoint, Communications of NII Shonan Meetings, Springer, January 2017 [DOI: 10.1007/978-981-10-4436-6], https://hal.inria.fr/hal-01615144.

# **Project-Team VISAGES**

# Vision, Action and information manaGement System in health

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

IN PARTNERSHIP WITH: CNRS INSERM Université Rennes 1

RESEARCH CENTER Rennes - Bretagne-Atlantique

THEME Computational Neuroscience and Medicine

## **Table of contents**

1.	Personnel	1005
2.	Overall Objectives	1007
3.	Research Program	1007
4.	Application Domains	1009
	4.1. Neuroimaging	1009
	4.2. Multiple sclerosis	1009
	4.3. Modeling of anatomical and anatomo-functional neurological patterns	1010
5.	Highlights of the Year	1010
6.	New Software and Platforms	1010
	6.1. Anima	1010
	6.2. autoMRI	1010
	6.3. MedInria	1011
	6.4. QtShanoir	1011
	6.5. Shanoir	1011
	6.6. ShanoirUploader	1012
	6.7. Platforms	1013
7.	New Results	1014
	7.1. Research axis 1: Medical Image Computing in Neuroimaging	1014
	7.1.1. Diffusion imaging	1014
	7.1.1.1. L2 Similarity Metrics for Diffusion Multi-Compartment Model Images Reg	istration
		1014
	7.1.1.2. Block-Matching Distortion Correction of Echo-Planar Images with Opposi	te Phase
	Encoding Directions	1014
	7.1.1.3. Diffusion MRI processing for multi-compartment characterization of brain	ı pathol-
	ogy	1015
	7.1.1.4. The challenge of mapping the human connectome based on diffusion tracto	graph <mark>ly015</mark>
	7.1.1.5. Comparison of inhomogeneity distortion correction methods in diffusion	MRI of
	the spinal cord	1015
	7.1.2. Arterial Spin Labeling:	1015
	7.1.2.1. Patch-based super-resolution for arterial spin labeling MRI	1016
	7.1.2.2. Resting-state functional ASL	1016
	7.1.2.3. Longitudinal atlas creation and brain development analysis	1016
	7.1.3. Quantitative relaxation times estimation and processing:	1017
	7.1.3.1. Gaining Insights Into Multiple Sclerosis Lesion Characteristics from Brai	n Tissue
	Microstructure Information: A Multi-Compartment T2 Relaxometry Approa	ch: 1017
	7.1.3.2. Multi-Compartment T2 Relaxometry Model Using Gamma Distribution R	epresen-
	tations: A Framework for Quantitative Estimation of Brain Tissue Microstr	uctures:
		1017
	7.1.4. Multi-modal EEG and fMRI Source Estimation Using Sparse Constraints:	1017
	7.2. Research axis 2: Applications in Neuroradiology and Neurological Disorders	1018
	7.2.1. Arterial Spin Labeling:	1018
	7.2.2. Hybrid EEG-fMRI Neurofeedback:	1018
	7.2.3. Multiple sclerosis:	1019
	7.2.4. Recovery imaging:	1019
	7.2.5. White matter connectivity analysis in patients suffering from depression:	1020
	7.2.6. Knowing and Remembering: Cognitive and Neural Influences of Familiarity on F	ecogni-
	tion Memory in Early Alzheimer's Disease (EPMR-MA):	1020
	7.2.7. Semantic Dementia Imaging:	1021
	7.3. Research axis 3: Management of Information in Neuroimaging	1021

8.	Bilateral Contracts and Grants with Industry	. 1021
	8.1. Bilateral Contracts with Industry	1021
	8.2. Bilateral Grants with Industry	1022
9.	Partnerships and Cooperations	. 1022
	9.1. Regional Initiatives	1022
	9.2. National Initiatives	1022
	9.2.1. Projet Fondation de France: PERINE	1022
	9.2.2. Projet Fondation de France: EPMR-MA	1023
	9.2.3. ANR "MAIA", 2015 generic projects program	1023
	9.2.4. Fondation pour la recherche médicale (FRM) - Project "Hybrid EEG/IRM Neurofeed	back
	for rehabilitation of brain pathologies	1023
	9.2.5. PHRC EMISEP: Evaluation of early spinal cord injury and late physical disabili	ty in
	Relapsing Remitting Multiple Sclerosis	1024
	9.2.6. Competitivity Clusters	1024
	9.2.6.1. The HEMISFER Project	1024
	9.2.6.2. France Life Imaging (FLI)	1025
	9.2.6.3. OFSEP	1025
	9.3. European Initiatives	1025
	9.4. International Initiatives	1026
	9.4.1. Inria Associate Teams Not Involved in an Inria International Lab	1026
	9.4.2. Inria International Partners	1027
	9.5. International Research Visitors	1027
	9.5.1. Visits of International Scientists	1027
	9.5.2. Visits to International Teams	1027
10.	Dissemination	. 1027
	10.1. Promoting Scientific Activities	1027
	10.1.1. Scientific Events Organisation	1027
	10.1.2. Scientific Events Selection	1027
	10.1.2.1. Member of the Conference Program Committees	1027
	10.1.2.2. Reviewer	1028
	10.1.3. Journal	1028
	10.1.3.1. Member of the editorial boards	1028
	10.1.3.2. Reviewing Activities	1028
	10.1.4. Leadership within the Scientific Community	1028
	10.1.5. Scientific Expertise	1028
	10.1.6. Invited Talks	1028
	10.2. Teaching - Supervision - Juries	1029
	10.2.1. Teaching	1029
	10.2.2. Supervision	1029
	10.2.3. Juries	1030
	10.3. Popularization	1030
	10.3.1. My thesis in 180 seconds	1030
	10.3.2. La semaine du cerveau (The Brain Week), from 13/03/2017 to 19/03/2017, Rennes	1030
	10.3.3. Journées Française de Radiologie 2017, 13-16 Octobre, Paris	1030
	10.3.4. 50 ans Inria, Paris	1030
	10.3.5. Journée du Président de la Société Française de Neurologie, Rennes, 7 Décembre 20	171030
11.	Bibliography	. 1030

## **Project-Team VISAGES**

Creation of the Project-Team: 2005 July 04

#### **Keywords:**

#### **Computer Science and Digital Science:**

- A3.1.2. Data management, quering and storage
- A3.1.3. Distributed data
- A3.1.7. Open data
- A3.1.8. Big data (production, storage, transfer)
- A3.2.4. Semantic Web
- A3.3.3. Big data analysis
- A3.4.1. Supervised learning
- A3.4.2. Unsupervised learning
- A3.4.3. Reinforcement learning
- A3.4.4. Optimization and learning
- A5.1.4. Brain-computer interfaces, physiological computing
- A5.2. Data visualization
- A5.3.2. Sparse modeling and image representation
- A5.3.3. Pattern recognition
- A5.3.4. Registration
- A5.4.1. Object recognition
- A5.4.6. Object localization
- A5.9.2. Estimation, modeling
- A6.2.3. Probabilistic methods
- A6.2.4. Statistical methods
- A6.3.3. Data processing
- A6.3.4. Model reduction
- A9.2. Machine learning
- A9.3. Signal analysis

#### **Other Research Topics and Application Domains:**

- B1.2. Neuroscience and cognitive science
- B1.2.1. Understanding and simulation of the brain and the nervous system
- B1.2.2. Cognitive science
- B2.1. Well being
- B2.2.2. Nervous system and endocrinology
- B2.2.6. Neurodegenerative diseases
- B2.5.1. Sensorimotor disabilities
- B2.5.2. Cognitive disabilities
- B2.6.1. Brain imaging

## **1. Personnel**

**Research Scientists** 

Christian Barillot [Team leader, CNRS, Senior Researcher, HDR] Emmanuel Caruyer [CNRS, Researcher] Olivier Commowick [Inria, Researcher] Camille Maumet [Inria, Researcher, from Nov 2017]

#### **Faculty Members**

Isabelle Bonan [Univ. Rennes Rennes I, Professor, HDR] Gilles Edan [Univ. Rennes, Professor] Jean-Christophe Ferré [Univ. Rennes Rennes, Professor, HDR] Jean-Yves Gauvrit [Univ. Rennes, Professor, HDR] Pierre Maurel [Univ. Rennes, Associate Professor]

#### **External Collaborators**

Jean-Marie Batail [Centre hospitalier régional et universitaire de Rennes] Florence Le Jeune [Univ. Rennes] Gabriel Robert [Univ. Rennes]

#### **Technical Staff**

Elise Bannier [Centre hospitalier régional et universitaire de Rennes] Yao Chi [Inria, granted by ANR] Isabelle Corouge [Univ. Rennes] Ines Fakhfakh [Inria, granted by ANR-OFSEP] Michael Kain [Inria, granted by ANR-FLI] Florent Leray [Inria, granted by ANR-FLI] Julien Louis [Inria, granted by ANR-FLI, until Mar. 2017] Aneta Morawin [Inria, granted by ANR-FLI] Mathieu Simon [Inria] Arnaud Touboulic [Inria, granted by ANR-OFSEP, from Feb. 2017]

#### **PhD Students**

Sudhanya Chatterjee [Univ. Rennes] Mathis Fleury [Inria, granted by FRM, from Nov. 2017] Renaud Hédouin [Inria, until Mar. 2017] Pierre-Yves Jonin [Centre hospitalier régional et universitaire de Rennes] Anne Kerbrat [Centre hospitalier régional et universitaire de Rennes] Antoine Legouhy [CNRS, granted by ANR and Brittany Region] Stéphanie Leplaideur [Centre hospitalier régional et universitaire de Rennes, from Oct. 2017] Cédric Meurée [Siemens Healthineers] Lorraine Perronnet [Inria, granted by Labex Cominlabs, until Apr. 2017] Maia Proisy [Centre hospitalier régional et universitaire de Rennes] Haykel Snoussi [Inria, granted by CHU] Corentin Vallée [Univ. Rennes]

#### **Post-Doctoral Fellows**

Quentin Duche [Inria, granted by INCR] Saman Noorzadeh [Inria, granted by Labex CominLabs, until Feb. 2017] Julie Coloigner [Inria, granted by INCR, from Apr. 2017] Benoit Combès [Inria, granted by CHU] Claire Cury [Inria, granted by Labex CominLabs, from Oct. 2017] Francesca Galassi [Inria, granted by INCR, from Apr. 2017] Giulia Lioi [Inria, granted by Labex Cominlabs, from Nov. 2017]

#### **Administrative Assistants**

Angélique Jarnoux [Inria, until February 2017] Armelle Mozziconacci [CNRS, from February 2017]

## 2. Overall Objectives

#### 2.1. Overall objectives

Medical Imaging, Neuroinformatics, Neuroimaging, Medical Image Computing, Modeling of normal and pathological behavior of the human brain, e-health & HealthGrids

The Unit/Project VISAGES U1228 is a research team jointly affiliated to INSERM (National Institute of Health and Scientific Research), Inria (National Institute of Research in Computer Sciences and Automation) and IRISA / UMR CNRS 6074, University of Rennes I. We are located in Rennes, France on both medical and sciences campus. The team was created in 2005. Our ambition is to set up a multidisciplinary team merging researchers in image processing and medical doctors. The goal of VISAGES is to constitute a multidisciplinary team. Even though, research in medical imaging could find motivation and recognition based on methodological breakthroughs alone, the ultimate goal, when dealing with medical imaging research, is to make the clinical practice benefit from the basic and applied research, while keeping the excellence of the methodological research. This objective entails the creation of teams encompassing clinical and scientific research team able to perform a research going from a novel and basic stage to original clinical experimentation with clear medical imaget.

Our research activities are focused on the research and development of new algorithms in medical imaging in the context of the pathologies of the central nervous system. In this context, we are addressing the general problems of the improved understanding of normal and pathological brain organs and systems behavior, at different scales, as well as the promotion and the support of Virtual Organizations of biomedical actors by means of healthgrid's technologies. The medical application objectives are focused on pathologies of the central nervous system, with a particular effort on extraction of new imaging biomarkers for brain pathologies (e.g., Multiple Sclerosis, neuropaediatrics, stroke, psychiatry, ...). More generally, our application objectives concern the following diseases: Multiple sclerosis, epilepsy, dementia, neuro-degenerative brain diseases, brain vascular diseases.

## 3. Research Program

#### 3.1. Research Program

The scientific foundations of our team concern the development of new processing algorithms in the field of medical image computing : image fusion (registration and visualization), image segmentation and analysis, management of image related information. Since this is a very large domain, which can be applied on numerous types of application; for seek of efficiency, the purpose of our methodological work primarily focuses on clinical aspects and for the most part on head and neck related diseases. In addition, we emphasize our research efforts on the neuroimaging domain. Concerning the scientific foundations, we have pushed our research efforts:

- In the field of image fusion and image registration (rigid and deformable transformations) with a special emphasis on new challenging registration issues, especially when statistical approaches based on joint histogram cannot be used or when the registration stage has to cope with loss or appearance of material (like in surgery or in tumor imaging for instance).
- In the field of image analysis and statistical modeling with a new focus on image feature and group analysis problems. A special attention was also to develop advanced frameworks for the construction of atlases and for automatic and supervised labeling of brain structures.
- In the field of image segmentation and structure recognition, with a special emphasis on the difficult problems of *i*) image restoration for new imaging sequences (new Magnetic Resonance Imaging protocols, 3D ultrasound sequences...), and *ii*) structure segmentation and labelling based on shape, multimodal and statistical information.

• Following past national projects where we had leading roles (e.g., Neurobase, NeuroLog, . . . ), we wanted to enhance the development of distributed and heterogeneous medical image processing systems.



Figure 1. The major overall scientific foundation of the team concerns the integration of data from the Imaging source to the patient at different scales: from the cellular or molecular level describing the structure and function, to the functional and structural level of brain structures and regions, to the population level for the modelling of group patterns and the learning of group or individual imaging markers.

As shown in Fig. 1, research activities of the VISAGES U1228 team are tightly coupling observations and models through integration of clinical and multi-scale data, phenotypes (cellular, molecular or structural patterns). We work on personalized models of central nervous system organs and pathologies, and intend to confront these models to clinical investigation studies for quantitative diagnosis, prevention of diseases, therapy planning and validation. These approaches are developed in a translational framework where the data integration process to build the models inherits from specific clinical studies, and where the models are assessed on prospective clinical trials for diagnosis and therapy planning. All of this research activity is conducted in tight links with the Neurinfo imaging platform environments and the engineering staff of the platform. In this context, some of our major challenges in this domain concern:

- The elaboration of new descriptors to study the brain structure and function (e.g., variation of brain perfusion with and without contrast agent, evolution in shape and size of an anatomical structure in relation with normal, pathological or functional patterns, computation of asymmetries from shapes and volumes).
- The integration of additional spatio-temporal imaging sequences covering a larger range of observation, from the molecular level to the organ through the cell (Arterial Spin Labeling, diffusion MRI, MR relaxometry, MR cell labeling imaging, PET molecular imaging, ...). This includes the elaboration of new image descriptors coming from spatio-temporal quantitative or contrast-enhanced MRI.
- The creation of computational models through data fusion of molecular, cellular, structural and functional image descriptors from group studies of normal and/or pathological subjects.
- The evaluation of these models on acute pathologies especially for the study of degenerative, psychiatric or developmental brain diseases (e.g., Multiple Sclerosis, Epilepsy, Parkinson, Dementia, Strokes, Depression, Schizophrenia, ...) in a translational framework.

In terms of methodological developments, we are particularly working on statistical methods for multidimensional image analysis, and feature selection and discovery, which include:
- The development of specific shape and appearance models, construction of atlases better adapted to a patient or a group of patients in order to better characterize the pathology;
- The development of advanced segmentation and modeling methods dealing with longitudinal and multidimensional data (vector or tensor fields), especially with the integration of new prior models to control the integration of multiscale data and aggregation of models;
- The development of new models and probabilistic methods to create water diffusion maps from MRI;
- The integration of machine learning procedures for classification and labeling of multidimensional features (from scalar to tensor fields and/or geometric features): pattern and rule inference and knowledge extraction are key techniques to help in the elaboration of knowledge in the complex domains we address;
- The development of new dimensionality reduction techniques for problems with massive data, which includes dictionary learning for sparse model discovery. Efficient techniques have still to be developed to properly extract from a raw mass of images derived data that are easier to analyze.

## 4. Application Domains

## 4.1. Neuroimaging

One research objective in neuroimaging is the construction of anatomical and functional cerebral maps under normal and pathological conditions. Many researches are currently performed to find correlations between anatomical structures, essentially sulci and gyri, where neuronal activation takes place, and cerebral functions, as assessed by recordings obtained by the means of various neuroimaging modalities, such as PET (Positron Emission Tomography), fMRI (Functional Magnetic Resonance Imaging), EEG (Electro-EncephaloGraphy) and MEG (Magneto-EncephaloGraphy). Then, a central problem inherent to the formation of such maps is to put together recordings obtained from different modalities and from different subjects. This mapping can be greatly facilitated by the use of MR anatomical brain scans with high spatial resolution that allows a proper visualization of fine anatomical structures (sulci and gyri). Recent improvements in image processing techniques, such as segmentation, registration, delineation of the cortical ribbon, modeling of anatomical structures and multi-modality fusion, make possible this ambitious goal in neuroimaging. This problem is very rich in terms of applications since both clinical and neuroscience applications share similar problems. Since this domain is very generic by nature, our major contributions are directed towards clinical needs even though our work can address some specific aspects related to the neuroscience domain.

## 4.2. Multiple sclerosis

Over the past years, a discrepancy became apparent between clinical Multiple sclerosis (MS) classification describing on the one hand MS according to four different disease courses and, on the other hand, the description of two different disease stages (an early inflammatory and a subsequently neurodegenerative phase). It is to be expected that neuroimaging will play a critical role to define in vivo those four different MS lesion patterns. An in vivo distinction between the four MS lesion patterns, and also between early and late stages of MS will have an important impact in the future for a better understanding of the natural history of MS and even more for the appropriate selection and monitoring of drug treatment in MS patients. MRI has a low specificity for defining in more detail the pathological changes which could discriminate between the different lesion types. However, it has a high sensitivity to detect focal and also widespread, diffuse pathology of the normal appearing white and gray matter. Our major objective within this application domain is then to define new neuroimaging markers for tracking the evolution of the pathology from high dimensional data (e.g., nD+t MRI) in the brain and the spinal cord. In addition, in order to complement MR neuroimaging data, we ambition to perform also cell labeling neuroimaging (e.g., MRI or PET) and to compare MR and PET data using standard and experimental MR contrast agents and radiolabeled PET tracers for activated microglia (e.g., USPIO or PK 11195). The goal is to define and develop, for routine purposes, cell specific and also quantitative imaging markers for the improved in vivo characterization of MS pathology.

## 4.3. Modeling of anatomical and anatomo-functional neurological patterns

The major objective within this application domain is to build anatomical and functional brain atlases in the context of functional mapping and for the study of developmental, neurodegenerative or even psychiatric brain diseases (Multiple sclerosis, Epilepsy, Parkinson, Dysphasia, Depression or even Alzheimer). This is a very competitive research domain; our contribution is based on our previous works in this field, and by continuing our local and wider collaborations. An additional objective within this application domain is to find new descriptors to study the brain anatomy and/or function (e.g., variation of brain perfusion, evolution in shape and size of an anatomical structure in relation with pathology or functional patterns, computation of asymmetries ...). This is also a very critical research domain, especially for many developmental or neurodegenerative brain diseases.

## 5. Highlights of the Year

## 5.1. Highlights of the Year

### 5.1.1. Recruitment

• Camille Maumet was recruited as Inria Researcher, starting from November 2017.

## 6. New Software and Platforms

## 6.1. Anima

KEYWORDS: Registration - Diffusion imaging - Medical imaging - Filtering - Relaxometry SCIENTIFIC DESCRIPTION: Anima is a set of libraries and tools developed by the team as a common repository of research algorithms. As of now, it contains tools for image registration, statistical analysis

(group comparison, patient to group comparison), diffusion imaging (model estimation, statistical analysis (group comparison, patient to group comparison), diffusion imaging (model estimation, tractography, etc.), quantitative MRI processing (quantitative relaxation times estimation, MR simulation), image denoising and filtering, and segmentation tools. All of these tools are based on stable libraries (ITK, VTK), making it simple to maintain.

- Participants: Aymeric Stamm, Fang Cao, Florent Leray, Guillaume Pasquier, Laurence Catanese, Olivier Commowick, Renaud Hedouin and René-Paul Debroize
- Contact: Olivier Commowick
- URL: https://github.com/Inria-Visages/Anima-Public/wiki

## 6.2. autoMRI

KEYWORDS: FMRI - MRI - ASL - FASL - SPM - Automation

SCIENTIFIC DESCRIPTION: This software is highly configurable in order to fit to a wide range of needs. Pre-processing includes segmentation of anatomical data, as well as co-registration, spatial normalization and atlas building of all data types. The analysis pipelines perform either within-group analysis or between-group or one subject-versus-group comparison and produce statistical maps of regions with significant differences. These pipelines can be applied to structural data to exhibit patterns of atrophy or lesions, to ASL (both pulsed or pseudo-continuous sequences) or PET data to detect perfusion or metabolic abnormalities, to relaxometry data to detect deviations from a template, to functional data - either BOLD or ASL - to outline brain activations related to block or event-related paradigms. In addition to the standard General Linear Model approach, the ASL pipelines implement an a contrario approach and, for patient-specific perfusion study, an heteroscedastic variance model. Besides, the vascular pipeline processes 4D MRA data and enables accurate assessment of hemodynamic patterns.

FUNCTIONAL DESCRIPTION: AutoMRI Based on MATLAB and the SPM8 toolbox, autoMRI provides complete pipelines to pre-process and analyze various types of images (anatomical, functional, perfusion, metabolic, relaxometry, vascular).

- Participants: Camille Maumet, Cédric Meurée, Elise Bannier, Fang Cao, Isabelle Corouge and Pierre Maurel
- Contact: Isabelle Corouge
- URL: http://www.irisa.fr/visages/

## 6.3. MedInria

KEYWORDS: Visualization - DWI - Health - Segmentation - Medical imaging

SCIENTIFIC DESCRIPTION: It aims at creating an easily extensible platform for the distribution of research algorithms developed at Inria for medical image processing. This project has been funded by the D2T (ADT MedInria-NT) in 2010, renewed in 2012. A fast-track ADT was awarded in 2017 to transition the software core to more recent dependencies and study the possibility of a consortium creation. The Visages team leads this Inria national project and participates in the development of the common core architecture and features of the software as well as in the development of specific plugins for the team's algorithm.

FUNCTIONAL DESCRIPTION: MedInria is a free software platform dedicated to medical data visualization and processing.

- Participants: Maxime Sermesant, Olivier Commowick and Théodore Papadopoulo
- Partners: HARVARD Medical School IHU LIRYC NIH
- Contact: Olivier Commowick
- URL: http://med.inria.fr

## 6.4. QtShanoir

KEYWORDS: Webservices - Soap - C++ - Health - DICOM - Plug-in - Medical imaging - Qt - Shanoir - Nifti SCIENTIFIC DESCRIPTION: QtShanoir is based on Qt/C++ librairie. It interacts with the Shanoir server using SOAP web services provided. This application queries the server and displays hierarchical data extracted in tree view. Data could also be easily downloaded or uploaded on the server. In order to extend the Shanoir environment, QtShanoir is developed to contain two shared libraries: - « GUI » that represents all user interfaces. - « DAO » that takes in charge the data model. This library assures the connection to the server and provides all QtShanoir services : research, download and upload of Processed Dataset (NIfTI). QtShanoir dynamic libraries are already reused and integrated in other projects: in the software medInria and in an under development command line program.

FUNCTIONAL DESCRIPTION: QtShanoir is a graphical client application of the medical imaging database Shanoir. This application provides various functionalities to satisfy researchers' needs. It allows users to: - explore neuroimaging data derived from multicenter research trials. Through an intuitive user interface, users could easily visualize voluminous amount of structured data: studies, patients and datasets extracted from Shanoir - download and to upload data from the server. This application is available on Windows, UNIX, MacOs X. It is integrated as a plugin in medInria, a multi-plateform for medical image processing and visualization.

- Participants: Alexandre Abadie, Guillaume Renard, Nicolas Wiest Daessle, Olivier Commowick and Wefa Hakem
- Contact: Christian Barillot
- URL: http://qtshanoir.gforge.inria.fr

## 6.5. Shanoir

SHAring NeurOImaging Resources

KEYWORDS: Neuroimaging - Medical imaging - PACS - Nifti - Data Sharing - DICOM - Health - Shanoir - Webservices - Data base - Biology - Web Application

FUNCTIONAL DESCRIPTION: SHAring NeurOImaging Resources (Shanoir, Previously InriaNeuroTk) is an open source software platform designed to structure, manage, archive, visualize and share neuroimaging data with an emphasis on multi-centric collaborative research projects. It provides common features of neuroimaging data management systems along with research-oriented data organization and enhanced accessibility.

Shanoir is a secured J2EE application running on a JBoss server, reachable via graphical interfaces in a browser or by third party programs via web services. It behaves as a repository of neuroimaging files coupled with a relational database hoding meta-data. The data model, based on OntoNeurolog, an ontology devoted to the neuroimaging field, is structured around research studies where of involved patients have examinations which either produce image acquisitions or clinical scores. Each image acquisition is composed of datasets represented by their acquisition parameters and image files. The system only keeps anonymous data.

Image files imports are possible from various sources (DICOM CDs, PACs, image files in NIfTI / Analyze format) using either online wizards, with completions of related meta-data, or commande line tools. Once de-identified during the import phase, DICOM header's customizable feature. Shanoir can also record any executed processing allowing to retrieve workflows applied to a particular dataset along with the intermediate data.

The clinical scores resulting from instrument based assessments (e.g. neuropsychological tests) can also be entered and easily retrieved and exported in different formats (Excel, CSV, Xml). Scores and image acquisitions are bound together which makes relationship analysis possible. The instrument database is scalable an new measures can be added in order to meet specific project needs, by use of intuitive graphical interfaces.

Using cross-data navigation and advanced search criteria, the users can quickly point to a subset of data of data to be downloaded. Client side applications have as well been developed to illustrate how to locally access and exploit data through the available web services. With regards to security, the system requires authentication and user rights are tunable for each hosted studies. A study responsible can thereby define the users allowed to see, download or import data into his study or simply make it public.

Shanoir serves neuroimaging researchers in organizing efficiently their studies while cooperating with other laboratories. By managing patient privacy, Shanoir allows the exploitation of clinical data in a research context. It is finally a handy solution to publish and share data with a broader community.

Shanoir integrates the enterprise search platform, Apache Solr, to provide the users a vast array of advanced features such as near real-time indexing and queries, full-text search, faceted navigation, autosuggestion and autocomplete.

- Participants: Adrien Férial, Anthony Baire, Bernard Gibaud, Christian Barillot, Guillaume Renard, Justine Guillaumont, Michael Kain and Yao Yao
- Partners: Université de Rennes 1 CNRS INSERM
- Contact: Christian Barillot
- URL: http://shanoir.gforge.inria.fr

## 6.6. ShanoirUploader

KEYWORDS: PACS - Medical imaging - Neuroimaging - DICOM - Health - Biology - Java - Webservices - Shanoir

SCIENTIFIC DESCRIPTION: ShanoirUploader is a desktop application on base of JavaWebStart (JWS). The application can be downloaded and installed using an internet browser. It interacts with a PACS to query and retrieve the data stored on it. After this ShanoirUploader sends the data to a Shanoir server instance in order to import these data. This application bypasses the situation, that in most of the clinical network infrastructures a server to server connection is complicated to set up between the PACS and a Shanoir server instance.

FUNCTIONAL DESCRIPTION: ShanoirUploader is a Java desktop application that transfers data securely between a PACS and a Shanoir server instance (e.g., within a hospital). It uses either a DICOM query/retrieve connection or a local CD/DVD access to search and access images from a local PACS or the local CD/DVD. After having retrieved the data, the DICOM files are locally anonymized and then uploaded to the Shanoir server. A possible integration of a hash creation application for patient identifiers is provided as well. The primary goals of that application are to enable mass data transfers between different remote server instances and therefore reduce the waiting time of the users, when importing data into Shanoir. Most of the time during import is spent with data transfers.

- Participants: Christian Barillot, Ines Fakhfakh, Justine Guillaumont, Michael Kain and Yao Yao
- Contact: Christian Barillot
- URL: http://shanoir.gforge.inria.fr

## 6.7. Platforms

## 6.7.1. The Neurinfo Platform

VisAGeS is the founding actor of an experimental research platform which was installed in August 2009 at the University Hospital of Rennes. The University of Rennes 1, Inria, Inserm for the academic side, and the University Hospital of Rennes and the Cancer Institute "Eugene Marquis" for the clinical side, are partners of this neuroinformatics platform called Neurinfo<sup>0</sup>. This platform has been supported under the "Contrat de Projets Etat-Région" (Christian Barillot is the PI) and has received a total amount of 4.01 M $\in$  for the period 2007–2014. European (FEDER), National (through Ministry of research, Inria, Inserm and ANR) and local councils (Brittany Region, Ille et Vilaine, and Rennes Metropole) have joined their effort to support this operation for a total amount of 4 010 k $\in$  (600 k $\in$  for the infrastructures, 2 850 k $\in$  for the equipments and 560 k $\in$  for the functioning). This application was set up through the Regional PIMATGI initiative coordinated by INSERM in Brittany (C. Roux). The overall PIMATGI initiative served for the financing of three distinct, but complementary, platforms: Neurinfo, TheraFONC as a technical platform dedicated to therapy guided by functional imaging especially in the oncology domain (Inserm U650 - LaTIM, Dir. Ch. Roux, Brest), and TherA-Image as a platform dedicated to image guided mini-invasive surgery and therapy especially in the domain of cardio-vascular diseases (U642 -LTSI, Dir. L. Senhadji, Rennes).

Concerning the Neurinfo Platform, the activity domain is a continuum between methodological and technological research built around specific clinical research projects. The ambition is to do innovation in science, technology and medical technology transfer for the implementation on the clinical field. On the medical field, the translational research domain mainly concerns medical imaging and more specifically the clinical neurosciences. Among them are multiple sclerosis, epilepsy, neurodegenerative, neurodevelopmental and psychiatric diseases, surgical procedures of brain lesions, neuro-oncology and radiotherapy planning. Beyond these CNS applications, the platform is also open to alternative applications. Neurinfo ambitions to support the emergence of research projects based on their level of innovation, their pluri-disciplinarity and their ability to foster collaborations between different actors (public and private research entities, different medical specialties, different scientific profiles).

In this context, a research 3T MRI system (Siemens Verio) was acquired in summer 2009 in order to develop the clinical research in the domain of morphological, functional, structural and cellular in-vivo imaging. In 2014 a new equipment for simultaneous recording of EEG and MRI images was acquired from Brain Product. In 2015, a mock scanner for experimental set-up was acquired as well as a new High Performance Computing environment made of one large computing cluster and a data center that is shared and operated by the Inria center at IRISA (UMR CNRS 6074). The computation cluster (240 cores) and the data center (up to 50 TB) are dedicated to host and process imaging data produced by the Neurinfo platform, but also by other research partners that share their protocols on the Neurinfo neuroinformatics system (currently more than 30 sites).

<sup>&</sup>lt;sup>0</sup>http://www.neurinfo.org

VisAGeS and its partners in the Neurinfo project are committed to use this new research platform for developing new regional, national and international collaborations around fundamental and applied clinical research projects dealing with in-vivo medical imaging.

In 2016, VisAGeS has been awarded by IBISA as a "Plateforme d'excellence".

In 2017, funding was collected to replace the 3T Siemens Verio MRI. A 3T Siemens Prisma will be installed early 2018.

## 7. New Results

## 7.1. Research axis 1: Medical Image Computing in Neuroimaging

Extraction and exploitation of complex imaging biomarkers involve an imaging processing workflow that can be quite complex. This goes from image physics and image acquisition, image processing for quality control and enhancement, image analysis for features extraction and image fusion up to the final application which intends to demonstrate the capability of the image processing workflow to issue sensitive and specific markers of a given pathology. In this context, our objectives in the recent period were directed toward 4 major methodological topics:

## 7.1.1. Diffusion imaging

## 7.1.1.1. L2 Similarity Metrics for Diffusion Multi-Compartment Model Images Registration

Participants: Renaud Hédouin, Olivier Commowick, Emmanuel Caruyer, Christian Barillot.

Diffusion multi-compartment models (MCM) allow for a fine and comprehensive study of the white matter microstructure. Non linear registration of MCM images may provide valuable information on the brain for example through population comparison. State-of-the-art MCM registration however relies on pairing-based similarity measures where the one-to-one mapping of MCM compartments is required. This approach leads to non differentiabilities or discontinuities, which may turn into poorer registration. Moreover, these measures are often specific to one MCM compartment model. We proposed [34] two new MCM similarity measures based on the space of square integrable functions, applied to MCM characteristic functions. These measures are pairing-free and agnostic to compartment types. We derived their analytic expressions for multi-tensor models and proposed a spherical approximation for more complex models. Evaluation was performed on synthetic deformations and inter-subject registration, demonstrating the robustness of the proposed measures.

7.1.1.2. Block-Matching Distortion Correction of Echo-Planar Images with Opposite Phase Encoding Directions Participants: Renaud Hédouin, Olivier Commowick, Élise Bannier, Christian Barillot.

By shortening the acquisition time of MRI, Echo Planar Imaging (EPI) enables the acquisition of a large number of images in a short time, compatible with clinical constraints as required for diffusion or functional MRI. However such images are subject to large, local distortions disrupting their correspondence with the underlying anatomy. The correction of those distortions is an open problem, especially in regions where large deformations occur. We have proposed a new block-matching registration method to perform EPI distortion correction based on the acquisition of two EPI with opposite phase encoding directions (PED). It relies on new transformations between blocks adapted to the EPI distortion model, and on an adapted optimization scheme to ensure an opposite symmetric transformation. We have produced qualitative and quantitative results of the block-matching correction using different metrics on a phantom dataset and on in-vivo data. We have shown the ability of the block-matching to robustly correct EPI distortion even in strongly affected areas. This work has been published in IEEE Transactions on Medical Imaging [21].

#### 7.1.1.3. Diffusion MRI processing for multi-compartment characterization of brain pathology Participants: Renaud Hédouin, Olivier Commowick, Christian Barillot.

Diffusion weighted imaging (DWI) is a specific type of MRI acquisition based on the direction of diffusion of the brain water molecules. It allows, through several acquisitions, to model the brain microstructure, as white matter, which is significantly smaller than the voxel-resolution. To acquire a large number of images in a clinical setting, very-fast acquisition techniques are required as single-shot imaging. However these acquisitions suffer locally large distortions. We have proposed a block-matching registration method based on the acquisition of images with opposite phase-encoding directions (PED). This technique specially designed for Echo-Planar Images (EPI) robustly correct images and provides a deformation field. This field is applicable to an entire DWI series from only one reversed EPI allowing distortion correction with a minimal acquisition time cost. This registration algorithm has been validated both on phantom and on in vivo data and is available in our source medical image processing toolbox Anima. From these diffusion images, we are able to construct multi-compartments models (MCM) which can represent complex brain microstructure. Doing registration, averaging and atlas creation on these MCM images is required to perform studies and statistic analyses. We propose a general method to interpolate MCM as a simplification problem based on spectral clustering. This technique, which is adaptable for any MCM, has been validated on both synthetic and real data. Then, from a registered dataset, we performed a patient to population analysis at a voxel-level computing statistics on MCM parameters. Specifically designed tractography can also be used to make analysis, following tracks, based on individual anisotropic compartments. All these tools are designed and used on real data and contribute to the search of biomakers for brain diseases such as multiple sclerosis.

#### 7.1.1.4. The challenge of mapping the human connectome based on diffusion tractography **Participant:** Emmanuel Caruyer.

Tractography based on non-invasive diffusion imaging is central to the study of human brain connectivity. To date, the approach has not been systematically validated in ground truth studies. Based on a simulated human brain data set with ground truth tracts, we organized an open international tractography challenge, which resulted in 96 distinct submissions from 20 research groups. Here, we report the encouraging finding that most state-of-the-art algorithms produce tractograms containing 90 percent of the ground truth bundles (to at least some extent). However, the same tractograms contain many more invalid than valid bundles, and half of these invalid bundles occur systematically across research groups. Taken together, our results demonstrate and confirm fundamental ambiguities inherent in tract reconstruction based on orientation information alone, which need to be considered when interpreting tractography and encourages innovation to address its current limitations [26].

## 7.1.1.5. Comparison of inhomogeneity distortion correction methods in diffusion MRI of the spinal cord **Participants:** Haykel Snoussi, Emmanuel Caruyer, Christian Barillot.

Diffusion MRI (dMRI) is a modality that describes the geometry of neural architecture. Diffusion images suffer from various artifacts originating from subject and physiological motion, eddy currents and B0-field inhomogeneity. These can severely affect image quality particularly in the spine region. However, strategies exist to correct these distortions, including co-registration, point spread function, phase field map and reversed gradient polarity method (RGPM). We evalute various correction methods using RGPM which provides best results. More precisely, we compared Voss plus two other recent methods: Topup (FSL) and HySCO (ACID/SPM). This work was presented at the ESMRMB conference [38].

#### 7.1.2. Arterial Spin Labeling:

Our contributions on this topic are illustrated in Fig. 2. Arterial Spin Labeling (ASL) enables measuring cerebral blood flow in MRI without injection of a contrast agent. Perfusion measured by ASL carries relevant information for patients suffering from pathologies associated with singular perfusion patterns.

However this technique suffers from drawbacks such as low signal to noise ratio and poor resolution.



Figure 2. Summary of the image processing workflow that allows the quantification of brain perfusion and detection of potential perfusion defect on patients or populations

## 7.1.2.1. Patch-based super-resolution for arterial spin labeling MRI

Participants: Cédric Meurée, Pierre Maurel, Christian Barillot.

In this context, our contributions focused on a super resolution approach to reduce the influence of Partial Volume Effects (PVE) and obtain images close to the ones that would be acquired at a high resolution, but in a shorter scan duration. PVE are an important limitation of arterial spin labeling (ASL) acquisitions, impacting the validity of quantitative cerebral blood flow (CBF) estimations. This work consists of a super-resolution algorithm, which includes information of high resolution (HR) structural images to reconstruct HR CBF maps from low resolution ASL series, without increasing the acquisition time. Compared with nearest neighbor, trilinear and 3rd order spline interpolations, the proposed algorithm is found to generate a CBF image closer to the one obtained with a reference HR ASL acquisition. CBF calculations can therefore be improved by using this algorithm, which reduces the PVE [36].

#### 7.1.2.2. Resting-state functional ASL

Participants: Corentin Vallée, Isabelle Corouge, Pierre Maurel, Christian Barillot.

We have started to work on resting-state functional ASL (rs-fASL). Rs-fASL in clinical daily practice and academic research stay discreet compared to resting-state BOLD. However, by giving direct access to cerebral blood flow maps, rs-fASL could lead to significant clinical subject scaled application as CBF can be considered as a biomarker in common neuropathology. As a new topic, we started by building a viable long sequence for rs-fASL. We take advantage of the long duration of the sequence to assess the link between overall quality of rs-fASL and duration of acquisition. To this end, we consider typical functional areas of the brain, and assess their quality compared to gold standards depending on the duration of acquisition. While some more work remain to be done, we tend to show there is an optimal duration of acquisition for rs-fASL. This work was submitted for the next ISMRM Conference.

#### 7.1.2.3. Longitudinal atlas creation and brain development analysis

Participants: Antoine Legouhy, Olivier Commowick, Christian Barillot.

The study of brain development provides insights in the normal trend of brain evolution and enables early detection of abnormalities. We propose a method to quantify growth in three arbitrary orthogonal directions of the brain through linear registration. We introduce a 9 degrees of freedom transformation that gives the

1017

opportunity to extract scaling factors describing brain growth along those directions by registering a data base of subjects in a common basis. We apply this framework to create a longitudinal curve of scaling ratios along fixed orthogonal directions from 0 to 16 years highlighting anisotropic brain development. In pediatric image analysis, the study of brain development provides insights in the normal trend of brain evolution and enables early detection of abnormalities. Tools like longitudinal atlases allow to compute statistics on populations, understand brain variability at different ages to highlight changes in growth, shape, structure etc. We experimented different methods to perform longitudinal atlases. This work was submitted for the next ISMRM Conference.

#### 7.1.3. Quantitative relaxation times estimation and processing:

The VisAGeS team has proposed new methodologies to exploit new relaxometry sequences, able to provide direct information on tissue properties (T1, T2, T2\* relaxation times) and their alteration in diseases. Such sequences have a great potential in diagnostic and evolution study of patients suffering from various neurological diseases.

7.1.3.1. Gaining Insights Into Multiple Sclerosis Lesion Characteristics from Brain Tissue Microstructure Information: A Multi-Compartment T2 Relaxometry Approach: Portiginante: Sudhamua Chattariaa Oliviar Commoniale Christian Parillet

Participants: Sudhanya Chatterjee, Olivier Commowick, Christian Barillot.

In addition to raw relaxation times, we have also studied other estimation methods able, from T2 relaxometry sequences, to estimate the fraction of myelin (myelin water fraction) inside each voxel, a quantity that may be largely impacted in neurological diseases. To this end, we have proposed new multi-compartment T2 estimation methods [42] with a new water three-compartment T2 model of tissue bounded water (free water, axons and cells, and myelin), using variable projection to make the estimation faster and more robust. Clinical trends and pathogenetic ways of onset and progression of Multiple Sclerosis (MS) in patients suggest that MS is a highly heterogeneous disease. MS is predominantly a White Matter (WM) disease, which is mainly composed of myelinated axons and neuroglia type cells. Demyelination and axonal loss characterize the condition of MS in a patient. However, they follow varying trends in patients. In this work, we propose a method in which T2 relaxometry data is used to obtain a quantitative brain tissue microstructure information. This information is then studied to check its corroborations with pathogenetic understanding of MS in literature [41].

7.1.3.2. Multi-Compartment T2 Relaxometry Model Using Gamma Distribution Representations: A Framework for Quantitative Estimation of Brain Tissue Microstructures:

Participants: Sudhanya Chatterjee, Olivier Commowick, Christian Barillot.

Advanced MRI techniques (e.g., d-MRI, MT, relaxometry etc.) can provide quantitative information of brain tissues. Image voxels are often heterogeneous in terms of microstructure information due to physical limitations and imaging resolution. Quantitative assessment of the brain tissue microstructure can provide valuable insights into neurodegenerative diseases (e.g., Multiple Sclerosis). In this work, we propose a multicompartment model for T2-Relaxometry to obtain brain microstructure information in a quantitative framework. The proposed method allows simultaneous estimation of the model parameters [42].

## 7.1.4. Multi-modal EEG and fMRI Source Estimation Using Sparse Constraints:

Participants: Saman Noorzadeh, Pierre Maurel, Christian Barillot.

In this work, a multi-modal approach is presented and validated on real data to estimate the brain neuronal sources based on EEG and fMRI. Combining these two modalities can lead to source estimations with high spatio-temporal resolution. The joint method is based on the idea of linear model already presented in the literature where each of the data modalities are first modeled linearly based on the sources. Afterwards, they are integrated in a joint framework which also considers the sparsity of sources. The sources are then estimated with the proximal algorithm. The results are validated on real data and show the efficiency of the joint model compared to the uni-modal ones. We also provide a calibration solution for the system and demonstrate the effect of the parameter values for uni- and multi-modal estimations on 8 subjects [37].

# 7.2. Research axis 2: Applications in Neuroradiology and Neurological Disorders

## 7.2.1. Arterial Spin Labeling:

Participants: Jean-Christophe Ferré, Maia Proisy, Isabelle Corouge, Élise Bannier, Christian Barillot.

Arterial Spin Labeling is an attractive perfusion MRI technique due to its complete non-invasiveness. However it still remains confidential in clinical practice. Over the years, we have developed several applications to evaluate its potential in different contexts. In 2017, in the context of the MALTA project, we focused on the application of ASL to activation-fMRI. Functional Arterial Spin Labeling (fASL) has demonstrated its greater specificity as a marker of neuronal activity than the reference BOLD fMRI for motor activation mapping in healthy volunteers. Motor fASL was yet to be investigated in the context of tumors, under the assumption that fASL would be less sensitive to venous contamination induced by the hemodynamics remodeling in the tumor vicinity than BOLD fMRI. As the arterial transit time may be shortened in activation areas, we explored the ability of fASL to map the motor areas at different post-labeling delays (PLD) in healthy subjects and patient with brain tumor. As part of the PhD of Maia Proisy, we have also been working on processing and analyse MR perfusion images using arterial spin labeling in neonates and children for several purposes:

- ASL and TOF-MRA are two totally non-invasive, easy-to-use MRI sequences for children in emergency settings. Hypoperfusion associated with homolateral vasospasm may suggest a diagnosis of migraine with aura (published in Cephalagia and presented in 3 congresses including RSNA)
- Investigation of brain perfusion evolution between 6 month and 15 years using ASL sequence in order to provide reference values in this age range (Measurement of pediatric regional cerebral blood flow from 6 months to 15 years of age article under revision, presented in one national congress)
- Work in Progress: ASL perfusion images in 20 neonates with hypoxic-ischemic encephalopathy that underwent MRI on day-of-life 3 and day-of-life 10.

#### 7.2.2. Hybrid EEG-fMRI Neurofeedback:

Participants: Lorraine Perronnet, Marsel Mano, Élise Bannier, Mathis Fleury, Giulia Lioi, Christian Barillot.

Over the last 4 years, we developed a whole new range of activities around hybrid EEG-MR imaging and neurofeedback for brain rehabilitation. We propose to combine advanced instrumental devices (Hybrid EEG and MRI platforms), with new man-machine interface paradigms (Brain computer interface and serious gaming) and new computational models (source separation, sparse representations and machine learning) to provide novel therapeutic and neuro-rehabilitation paradigms in some of the major neurological and psychiatric disorders of the developmental and the aging brain. We first performed a thorough state-of-the-art of Neurofeedback (NF) and restorative Brain Computer Interfaces (BCI) under EEG and fMRI modality as well as of EEG-fMRI integration, with a particular focus on applications in depression and motor rehabilitation. This enabled us to design a NF protocol based on motor imagery and compatible with EEG and fMRI. We implemented different types of feedback and compared for the first time the effects of unimodal EEG-NF and fMRI-NF versus bimodal EEG-fMRI-NF by looking both at EEG and fMRI activations. We also introduced a new feedback metaphor for bimodal EEG-fMRI-neurofeedback that integrates both EEG and fMRI signal in a single bi-dimensional feedback (a ball moving in 2D). The participants to this study were able to regulate activity in their motor regions in all NF conditions. Our results also suggest that that EEG-fMRI-neurofeedback could be more specific or more engaging than EEG-NF alone [31].

All the experiments were performed on the Neurinfo platform which is equipped with an EEG MR compatible 64-channel device in 2014 to perform joint EEG and BOLD or ASL fMRI. We developed, installed and successfully tested a hybrid EEG-fMRI platform for bimodal NF experiments. Our system is based on the integration and the synchronization of an MR-compatible EEG and fMRI acquisition subsystems. We developed two real-time pipelines for EEG and fMRI that handle all the necessary signal processing, the joint NF block that calculates and fuses the NF and a visualization block that displays the NF to the subject. The control and the synchronization of both subsystems with each other and with the experimental protocol

is handled by the NF Control. Our platform showed very good real-time performance with various preprocessing, filtering, and NF estimation and visualization methods. Its modular architecture is easily adaptable to different experimental environments, and offers high efficiency for optimal real-time NF applications [27].

These developments came as part of the HEMISFER project which is conducted through a very complementary set of competences over the different teams involved (Visages Inserm U1228, HYBRID and PANAMA Teams from Inria/IRISA, EA 4712 team from University of Rennes I and ATHENA team from Inria Sophia-Antipolis). The overall principle of this project is illustrated in Fig. 3.



Figure 3. Principle of the Hybrid EEG:fMRI environement set up and used by the HEMISFER project

#### 7.2.3. Multiple sclerosis:

**Participants:** Anne Kerbrat, Gilles Edan, Jean-Christophe Ferré, Benoit Combès, Olivier Commowick, Élise Bannier, Sudhanya Chatterjee, Haykel Snoussi, Emmanuel Caruyer, Christian Barillot.

The VisAGeS research team has a strong focus on applying the developed methodologies (illustrated in research axis 1) to multiple sclerosis (MS) understanding and the prediction of its evolution. Related to the EMISEP project on spinal cord injury evolution in MS, a first work investigated the magnetization transfer reproducibility across centers in the spinal cord and was accepted for presentation at ESMRMB [33]. Based on this work, a second work investigated the sensitivity of magnetization transfer to assess diffuse and focal burden in MS patients [43]. In parallel, methodological developments have addressed spinal cord diffusion data analysis, starting with a comparaison of several distortion correction methods [38].

Finally, we investigated myelin water fraction (MWF) estimation on multiple sclerosis and demonstrated in longitudinal studies [41] how these figures can be related with lesion evolution, paving the way towards myelin oriented MS evaluation of patient future evolution prediction (and thus treatment adaptation) and joint studies between different quantitative imaging modalities (e.g., diffusion).

#### 7.2.4. Recovery imaging:

Participants: Isabelle Bonan, Stephanie Leplaideur, Élise Bannier, Jean-Christophe Ferré, Christian Barillot.

More common after a right hemispheric brain injury, misperception of body in space, impacting moves and posture is often associated with disturbance of spatial attention (behavioural symptoms of a failure in spontaneously reorienting attention to stimulus information in the left field). While different subjects use different references in their elaboration of spatial representation, body-centered coordinate systems are the most prevalent. As part of an fMRI substudy of a national research study on balance disorder rehabilitation, we investigated differences in activations during body-centered spatial tasks in corporeal and in extracorporeal space. Healthy controls and stroke patients were included in this fMRI sub study comprising 2 egocentric spatial tasks: perception of the midsagittal plane in extracorporeal space (straight-ahead task) and in corporeal space (longitudinal axis task). Results obtained on healthy control data were presented at the SOFMER conference and the journal paper is under review. For both tasks, cerebral activations largely dominated in the right hemisphere and essentially involved the right frontoparietal network. In addition, the straightahead task presented specific activations in the temporoparieto-insular cortex and thalamic areas. Patient data processing is ongoing in the context of an MD-PhD. In parallel, a master study investigated the brain structural connections between the cortical areas obtained from the fMRI study using diffusion MRI and the white matter query language.

#### 7.2.5. White matter connectivity analysis in patients suffering from depression:

Participants: Julie Coloigner, Jean-Marie Batail, Jean-Christophe Ferré, Isabelle Corouge, Christian Barillot.

The mood depressive disorder (MDD) is a common chronically psychiatric disorder with an estimated lifetime prevalence reported to range from 10 percent to 15 percent worldwide. This disease is characterized by an intense dysregulation of affect and mood as well as additional abnormalities including cognitive dysfunction, insomnia, fatigue and appetite disturbance. Despite the extensive therapy options available for depression, up to 80 percent of patients will suffer from a relapse [1]. Consequently, exhibiting imaging biomarkers of this disease will support both a better understanding of the neural correlates underlying the depression, and a better diagnosis and treatment of individual depressed patients. Previous studies of structural and functional magnetic resonance imaging have reported several microstructural abnormalities in the prefrontal cortex, anterior cingulate cortex, hippocampus and thalamus [2]. These observations suggest a dysfunction of the circuits connecting frontal and subcortical brain regions, leading to a "disconnection syndrome" [3]. Given the small sample size used in the past studies, we proposed a more robust analysis using a larger cohort of patients suffering from depression, called LONGIDEP. The latter is a routine care cohort of patients suffering from mood depressive disorder who underwent a clinical evaluation, neuropsychological testing and brain MRI. The population sample consists of 125 patients suffering from depression and 65 healthy age and gendermatched, control subjects. A composite measure of medication load for each patient was assessed using a previously established method [4]. We investigated alterations of white matter integrity using a voxel-based analysis based on fractional anisotropy (FA) and the apparent diffusion coefficient (ADC) in patients with depression. Using graph theory-based analysis, we also examined white matter changes in the organization of networks in patients suffering from depression. Our findings provide robust evidence that the reduction of white-matter integrity in the interhemispheric connections and fronto-limbic neuronal circuits may play an important role in MDD pathogenesis. These results are consistent with an overall hypothesis that depression involves a disconnection of prefrontal, striatal, and limbic emotional areas.

# 7.2.6. Knowing and Remembering: Cognitive and Neural Influences of Familiarity on Recognition Memory in Early Alzheimer's Disease (EPMR-MA):

Participants: Pierre-Yves Jonin, Quentin Duché, Élise Bannier, Christian Barillot.

Inclusion of the 20 healthy participants in the "EPMR-MA" study (clinical trials ID NCT02492529) has been achieved, the inclusion phase will be achieved before 30th, december, 2017. Healthy controls data are preprocessed and the first analysis workflow proved promising, it should allow submitting a first paper at the beginning of 2018.

#### 7.2.7. Semantic Dementia Imaging:

Participants: Jean-Christophe Ferré, Isabelle Corouge, Elise Bannier, Christian Barillot.

After demonstrating the relative preservation of fruit and vegetable knowledge in patients with semantic dementia (SD), we sought to identify the neural substrate of this unusual category effect. Nineteen patients with SD performed a semantic sorting task and underwent a morphometric 3T MRI scan. The grey-matter volumes of five regions within the temporal lobe were bilaterally computed, as well as those of two recently described areas (FG1 and FG2) within the posterior fusiform gyrus. In contrast to the other semantic categories we tested, fruit and vegetable scores were only predicted by left FG1 volume. We therefore found a specific relationship between the volume of a subregion within the left posterior fusiform gyrus and performance on fruits and vegetables in SD. We argue that the left FG1 is a convergence zone for the features that might be critical to successfully sort fruits and vegetables. We also discuss evidence for a functional specialization of the fusiform gyrus along two axes (lateral medial and longitudinal), depending on the nature of the concepts and on the level of processing complexity required by the ongoing task [28].

## 7.3. Research axis 3: Management of Information in Neuroimaging

**Participants:** Élise Bannier, Christian Barillot, Yao Chi, Isabelle Corouge, Olivier Commowick, Inès Fakhfakh, Michael Kain, Florent Leray, Julien Louis, Aneta Morawin, Mathieu Simon, Arnaud Touboulic.

The major topic that has been reached in the period concerns the sharing of data and processing tools in neuroimaging (through the ANR Neurolog and VIP projects, and more recently the "Programme d'Investissement d'Avenir" project such as OFSEP and FLI-IAM) that led to build a suitable architecture to share images and processing tools). Our overall goal within these projects was to set up a computational infrastructure to facilitate the sharing of neuroimaging data, as well as image processing tools, in a distributed and heterogeneous environment. These consortiums gathered expertises coming from several complementary domains: image processing in neuroimaging, workflows and grid computing, ontology development and ontology-based mediation. Shanoir (SHAring NeurOImaging Resources) is one of the major outcome of these projects. Shanoir uses semantics for concepts organization that are defined by the OntoNeuroLOG ontology. OntoNeuroLOG reuses and extends the OntoNeuroBase ontology. Both were designed using the same methodological framework, based on the use of the foundational ontology DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering), and the use of a number of core ontologies, that provide generic, basic and minimal concepts and relations in specific domains such as Artefacts. Shanoir aims at establishing the conditions allowing, through the Internet, to share distributed information sources in neuroimaging, whether these sources are located in various centers of experimentation, clinical departments of neurology, or research centers in cognitive neurosciences or image processing. This enables a large variety of users to diffuse, exchange or reach neuroimaging information with appropriate access means, in order to be able to retrieve information almost as easily as if the data were stored locally by means of the "cloud computing" Storage as a Service (SaaS) concept. The Shanoir environment has been succesfully deployed to the Neurinfo platform were it is routinely used to manage images of the research studies. It is also currently being deployed for two large projets: OF-SEP ("Observatoire Français de la Sclérose en Plaques") where up to 30000 patients will be acquired on a ten years frame, and the Image Analysis and Management (IAM) node of the France Life Imaging national infrastructure (FLI-IAM). Our team VisAGeS fulfills multiple roles in this nation-wide FLI project. Christian Barillot is the chair of the IAM node, Olivier Commowick is participating in the working group workflow and image processing and Michael Kain is the technical manager of the node. Apart from the team members, software solutions like MedInria and Shanoir are part of the final infrastructure software solutions.

## 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Contracts with Industry

8.1.1. Siemens

In the context of the Neurinfo imaging platform, a master research agreement between Siemens SAS -Healthcare and University of Rennes 1 was signed in October 2011 for 5 years and renewed in 2016. This contract defines the terms of the collaboration between Siemens, Visages and the Neurinfo platform. From this research agreement contract, Neurinfo has received work in progress (WIP) sequences from Siemens in the form of object code for evaluation in the context of clinical research. The Neurinfo platform has also received source code of selected MRI sequences. As an example, the diffusion sequence code was modified to load arbitrary diffusion gradient waveforms for the FastMicroDiff project led by E. Caruyer. This is crucial in the collaboration since it enables the development of MRI sequences on site. Siemens currently provides research resources through the funding of a PhD student (Cédric Meurée: CIFRE Inria / Siemens grant).

## 8.2. Bilateral Grants with Industry

The PhD of Cédric Meurée is funded by Siemens Healthineers under a CIFRE grant.

## 9. Partnerships and Cooperations

## 9.1. Regional Initiatives

## 9.1.1. Allocation d'Installation Scientifique – Rennes Métropole

Participant: Emmanuel Caruyer.

Diffusion MRI has been a tremendous tool for the diagnosis of a number of brain pathologies such as abnormal development, neuro-degenerative or inflammatory disorders or brain tumors. Typical resolution in diffusion MRI is about 2mm – this suggests that in white matter, any volume element may contain millions of axons. Although currently we can characterize molecular diffusion, recent developments in diffusion MRI have shown the possibility to quantify more specifically some physical tissue parameters in white matter, such as axonal density and diameter: this means that we can retrieve information from a much smaller scale than the typical imaging resolution.

Acquisition time for this kind of measurements remains long and largely incompatible with in vivo application in humans. This projects aims at developing novel signal processing and acquisition methods for the reconstruction of microstructural informations in a reasonable acquisition time. We will study how sparse representations can be applied to the diffusion signal, in order to enable microstructure information reconstruction. In conjunction with this, we will develop acquisition sequences adapted to these sparse representations, in order to reconstruct the diffusion signal from fewer measurements, using results from the compressive sensing theory.

## 9.2. National Initiatives

#### 9.2.1. Projet Fondation de France: PERINE

**Participants:** Élise Bannier, Isabelle Corouge, Julie Coloigner, Maia Proisy, Jean-Christophe Ferré, Christian Barillot.

This study evaluates the effect of prenatal exposure to neurotoxicants on the developing brain. Following previous studies in the PELAGIE cohort this MRI study involves ASL, Diffusion and working memory as well as motor inhibition BOLD fMRI together with neuropsychological tests in children. Inclusions have started in November 2014 and lasted for 2 years. The MRI acquisitions of the PERINE projects have all been performed and 101 children included. A PhD started in January 2017 to process the functional MRI data of this project and Julie Coloigner was hired as a post doc to work on the Diffusion and ASL data.

Participants: Pierre-Yves Jonin, Élise Bannier, Christian Barillot, Quentin Duché.

This project evaluates memory effects in healthy adults and in patients presenting cognitive impairments using BOLD fMRI and diffusion MRI. The inclusions of patients started in 2016 and all inclusions will be over by the end of 2017. Quentin Duché was hired to process the functional MRI and diffusion data end of 2016 and his contract was extended until May 2018.

#### 9.2.3. ANR "MAIA", 2015 generic projects program

**Participants:** Maia Proisy, Pierre Maurel, Antoine Legouhy, Olivier Commowick, Isabelle Corouge, Jean-Christophe Ferré, Christian Barillot.

Each year in France, 55 000 children are born prematurely, i.e., before the 37th week of gestation. Long-term studies of the outcome of prematurely born infants have clearly documented that the majority of such infants may have significant motor, cognitive, and behavioral deficits.

However, there is a limited understanding of the nature of the cerebral abnormality underlying these adverse neurologic outcomes. In this context, the emergence of new modalities of 3D functional MRI, e.g., Arterial Spin Labeling (ASL), or optical imaging technologies, e.g., Near InfraRed Spectroscopy (NIRS), brings new perspectives for extracting cognitive information, via metabolic activity measures. Other classical techniques devoted to cerebral signal measurement, such as ElectroEncephaloGraphy (EEG), provide cognitive information at the cortical level. Each of these various non-invasive imaging technologies brings substantial and specific information for the understanding of newborn brain development.

This project aims at developing innovative approaches for multi-image / multi-signal analysis, in order to improve neurodevelopment understanding methods. From a fundamental point of view, mathematics and computer science have to be considered in association with imaging physics and medicine, to deal with open issues of signal and image analysis from heterogeneous data (image, signal), considered in the multiphysics contexts related to data acquisition (magnetic, optic, electric signals) and biophysics modeling of the newborn brain. A sustained synergy between all these scientific domains is then necessary.

Finally, the sine qua non condition to reach a better understanding of the coupled morphological- cognitive development of premature newborns, is the development of effective software tools, and their distribution to the whole medical community. The very target of this project will be the design of such software tools for medical image / signal analysis, actually operational in clinical routine, and freely available. Academic researchers and industrial partners will work in close collaboration to reach that ambitious goa.

## 9.2.4. Fondation pour la recherche médicale (FRM) - Project ''Hybrid EEG/IRM Neurofeedback for rehabilitation of brain pathologies

**Participants:** Élise Bannier, Jean-Marie Batail, Isabelle Bonan, Isabelle Corouge, Jean-Christophe Ferré, Jean-Yves Gauvrit, Pierre Maurel, Mathis Fleury, Giulia Lioi, Christian Barillot.

The goal of this project is to make full use of neurofeedback (NF) paradigm in the context of brain rehabilitation. The major breakthrough will come from the coupling associating functional and metabolic information from Magnetic Resonance Imaging (fMRI) to Electro-encephalography (EEG) to "optimize" the neurofeedback protocol. We propose to combine advanced instrumental devices (Hybrid EEG and MRI platforms), with new hybrid Brain computer interface (BCI) paradigms and new computational models to provide novel therapeutic and neuro-rehabilitation paradigms in some of the major mental and neurological disorders of the developmental and the aging brain (stroke, language disorders, Mood Depressive Disorder (MDD), ...). Though the concept of using neurofeedback paradigms for brain therapy has somehow been experimented recently (mostly through case studies), performing neurofeedback through simultaneous fMRI and EEG has almost never been done before so far (two teams in the world including us within the HEMISFER CominLabs project). This project will be conducted through a very complementary set of competences over the different involved teams: VISAGES U1228, HYBRID and PANAMA Teams from Inria/Irisa Rennes and EA 4712 team from U. of Rennes I.

## 9.2.5. PHRC EMISEP: Evaluation of early spinal cord injury and late physical disability in Relapsing Remitting Multiple Sclerosis

**Participants:** Élise Bannier, Christian Barillot, Emmanuel Caruyer, Benoit Combès, Olivier Commowick, Gilles Edan, Jean-Christophe Ferré, Anne Kerbrat, Haykel Snoussi.

Multiple Sclerosis (MS) is the most frequent acquired neurological disease affecting young adults (1/1000 inhabitants in France) and leading to impairment. Early and well adapted treatment is essential in patients presenting aggressive forms of MS. This PHRC project focusses on physical impairment and especially on the ability to walk. Several studies, whether epidemiologic or based on brain MRI, have shown that several factors were likely to announce aggressive development of the disease, such as age, number of focal lesions on baseline MRI, clinical activity. However, these factors only partially explain physical impairment progression, preventing their use at the individual level. Spinal cord is often affected in MS, as demonstrated in postmortem or imaging studies. Yet, early radiological depiction of spinal cord lesions is not always correlated with clinical symptoms. Preliminary data, on reduced number of patients, and only investigating the cervical spinal cord have shown that diffuse spinal cord injury, observed via diffusion or magnetisation transfer imaging, would be correlated with physical impairment as evaluated by the EDSS score. Besides, the role of early spinal cord affection (first two years) in the evolution of physical impairment remains unknown.

In this project, we propose to address these different issues and perform a longitudinal study on Relapsing Remitting Multiple Sclerosis (RRMS) patients, recruited in the first year of the disease. Our goal is to show that diffuse and focal lesions detected spinal cord MRI in the first 2 years can be used to predict disease evolution and physical impairment at 5 years. Twelve centers are involved in the study to include 80 patients.

To date, all subjects have been included. H. Snoussi is working in the scope of his PhD thesis on diffusion imaging in the spinal cord starting with distortion correction. The results of this study were presented at the ESMRMB 2017 conference [38].

B. Combès started as a post doc in November 2016 to process the EMISEP imaging data, starting with morphological data processing (registration, segmentation) and magnetization transfer data processing. Preliminary results were presented at the ESMRMB and ECTRIMS 2017 conferences [33] [43].

## 9.2.6. Competitivity Clusters

#### 9.2.6.1. The HEMISFER Project

**Participants:** Élise Bannier, Jean-Marie Batail, Isabelle Bonan, Isabelle Corouge, Claire Cury, Jean-Christophe Ferré, Jean-Yves Gauvrit, Marsel Mano, Pierre Maurel, Saman Norzade, Lorraine Perronnet, Christian Barillot.

The HEMISFER project ("Hybrid Eeg-MrI and Simultaneous neuro-FEedback for brain Rehabilitation") will be conducted at Inria Rennes with the support of the Cluster of Excellence "CominLabs"<sup>0</sup>. The goal of HEMISFER is to make full use of the neurofeedback paradigm in the context of rehabilitation and psychiatric disorders. The major breakthrough will come from the use of a coupling model associating functional and metabolic information from Magnetic Resonance Imaging (fMRI) to Electro-encephalography (EEG) to "enhance" the neurofeedback protocol. We propose to combine advanced instrumental devices (Hybrid EEG and MRI platforms), with new man-machine interface paradigms (Brain computer interface and serious gaming) and new computational models (source separation, sparse representations and machine learning) to provide novel therapeutic and neuro-rehabilitation paradigms in some of the major neurological and psychiatric disorders of the developmental and the aging brain (stroke, attention-deficit disorder, language disorders, treatment-resistant mood disorders, ...). This project will be conducted with the HYBRID and PANAMA Teams from Inria Rennes, the EA 4712 team from University of Rennes I and the ATHENA team from Inria Sophia-Antipolis. This work will benefit from the research 3T MRI and MRI-compatible EEG systems provided by the NeurInfo in-vivo neuroimaging platform on which these new research protocols will be set up. A budget of 500keuros will be provided by the CominLabs cluster in the next 3 years to support this project (through experimental designs, PhDs, Post-docs and Expert Engineers).

<sup>&</sup>lt;sup>0</sup>https://iww.inria.fr/cominlabs-newsletter/april-2013-four-projects-selected/#hemisfer

#### 9.2.6.2. France Life Imaging (FLI)

**Participants:** Christian Barillot, Olivier Commowick, Michael Kain, Florent Leray, Julien Louis, Aneta Morawin, Mathieu Simon, Yao Chi.

France Life Imaging (FLI) is a proposed large-scale research infrastructure project aimed at establishing a coordinated and harmonized network of biomedical imaging in France. This project was recently selected by the call "Investissements d'Avenir - Infrastructure en Biologie et Santé". One node of this project is the node Information Analysis and Management (IAM), a transversal node build by a consortium of teams that will contribute to the construction of a network for data storage and information processing. Instead of building yet other dedicated facilities, the IAM node will use already existing data storage and information processing facilities (LaTIM Brest; CREATIS Lyon; CIC-IT Nancy; VisAGeS U1228 Inria Rennes; CATI CEA Saclay; LSIIT/ICube Strasbourg) that will increase their capacities for the FLI infrastructure. Inter-connections and access to services will be achieved through a dedicated software platform that will be developed based on the expertise gained through successful existing developments. The IAM node has several goals. It aims first at building a versatile facility for data management that will inter-connect the data production sites and data processing for which state-of-the-art solutions, hardware and software, will be available to infrastructure users. Modular solutions are preferred to accommodate the large variety of modalities acquisitions, scientific problems, data size, and adapted for future challenges. Second, it aims at offering the latest development that will be made available to image processing research teams. The team VisAGeS fulfills multiple roles in this nation-wide project. Christian Barillot is the chair of the node IAM, Olivier Commowick is participating in the working group workflow and image processing and Michael Kain the technical manager. Apart from the team members, software solutions like MedInria and Shanoir will be part of the final software platform.

#### 9.2.6.3. OFSEP

**Participants:** Élise Bannier, Christian Barillot, Olivier Commowick, Gilles Edan, Jean-Christophe Ferré, Michael Kain, Inès Fakhfakh.

The French Observatory of Multiple Sclerosis (OFSEP) is one of 10 projects selected in January 2011 in response to the call for proposal in the "Investissements d'Avenir - Cohorts 2010" program launched by the French Government. It allows support from the National Agency for Research (ANR) of approximately  $\in$  10 million for 10 years. It is coordinated by the Department of Neurology at the Neurological Hospital Pierre Wertheimer in Lyon (Professor Christian Confavreux), and it is supported by the EDMUS Foundation against multiple sclerosis, the University Claude Bernard Lyon 1 and the Hospices Civils de Lyon. OFSEP is based on a network of neurologists and radiologists distributed throughout the French territory and linked to 61 centers. OFSEP national cohort includes more than 50,000 people with Multiple Sclerosis, approximately half of the patients residing in France. The generalization of longitudinal monitoring and systematic association of clinical data and neuroimaging data is one of the objectives of OFSEP in order to improve the quality, efficiency and safety of care and promote clinical, basic and translational research in MS. For the concern of data management, the Shanoir platform of Inria has been retained to manage the imaging data of the National OFSEP cohort in multiple sclerosis.

## 9.3. European Initiatives

#### 9.3.1. Collaborations in European Programs, Except FP7 & H2020

#### • OpenAire-Connect

The OpenAire-Connect H2020 project will introduce and implement the concept of Open Science as a Service (OSaaS) on top of the existing OpenAIRE infrastructure, delivering out-of-the-box, ondemand deployable tools. OpenAIRE-Connect will adopt an end-user driven approach (via the involvement of 5 prominent research communities), and enrich the portfolio of OpenAIRE infrastructure production services with a Research Community Dashboard Service and a Catch-All Notification Broker Service. The first will offer publishing, interlinking, packaging functionalities to enable them to share and re-use their research artifacts (introducing methods, e.g., data,software, protocols). This effort, supported by the harvesting and mining "intelligence" of the OpenAIRE infrastructure, will provide communities with the content and tools they need to effectively evaluate and reproduce science. OpenAIRE-Connect will combine dissemination and training with OpenAIRE's powerful NOAD network engaging research communities and content providers in adopting such services. These combined actions will bring immediate and long-term benefits to scholarly communication stakeholders by affecting the way research results are disseminated, exchanged, evaluated, and re-used. In this project VisAGeS is acting, through CNRS, as the French coordinator to develop the link with the Neuroimaging research community. This will be performed in the context of the FLI-IAM national infrastructure.

- Participants: Christian Barillot; Michael Kain; Camille Maumet
- Partners: PI: CNR, Italy; Athena Research And Innovation Center In Information Communication & Knowledge Technologies, Greece; Uniwersytet Warszawski, Poland; JISC LBG, UK; Universitaet Bremen, Germany; Universidade Do Minho, Portugal; CNRS (Visages, Creatis), France; Universita Di Firenze, Italy; Institut De Recherche Pour Le Developpement (IRD), France; European Organization For Nuclear Research (CERN), Switzerland; International Center For Research On The Environment And The Economy, Greece
- Budget:  $2M \in (120k \in \text{ for CNRS})$

#### • Health

EIT Health aims to promote entrepreneurship and develop innovations in healthy living and active ageing, providing Europe with new opportunities and resources. EIT Health will enable citizens to lead healthier and more productive lives by delivering products, services and concepts that will improve quality of life and contribute to the sustainability of healthcare across Europe. EIT Health is a strong, diverse and balanced partnership of best-in-class organisations in education, research, technology, business creation and corporate and social innovation. EIT Health intends to foster cooperation and unlock Europe's innovation and growth potential – developing and retaining the best talents, creating high-quality jobs and boosting the global competitiveness of European industry. VisAGeS is involved in this project through the Inserm and Inria institutions. Christian Barillot is representing Inria as one expert in the dedicated WG "Healthy Brain". VisAGeS is also concerned by the WG "big data".

- Participants: Christian Barillot, Michael Kain
- Partners: see https://www.eithealth.eu/partners

## 9.4. International Initiatives

#### 9.4.1. Inria Associate Teams Not Involved in an Inria International Lab

#### 9.4.1.1. BARBANT

Title: Boston and Rennes, a Brain image Analysis Team

International Partner (Institution - Laboratory - Researcher):

Harvard University (United States) - Mathematics Department - Simon K. Warfield

Start year: 2015

See also: https://team.inria.fr/barbant/

BARBANT is an Inria associate team shared between Inria VisAGeS research team and the Computational Radiology Laboratory at the Boston Children's hospital (Harvard Medical School). This associate team aims at better understanding the behavior of normal and pathological Central Nervous System (CNS) organs and systems. Pathologies of particular interest to us are multiple sclerosis, psychiatric, and pediatric diseases such as pediatric multiple sclerosis or tuberous sclerosis. A major challenge is to characterize the future course of the pathological processes in each patient as early as possible in order to predict the progression of the disease and/or adverse neurological outcomes, and to develop better techniques for both monitoring response to therapy and for altering therapy (duration, dose and nature) in response to patient-specific changes in imaging characteristics.

At term, this project will allow to introduce objective figures to correlate qualitative and quantitative phenotypic markers coming from the clinic and image analysis, mostly at the early stage of the pathologies. This will allow for the selection or adaptation of the treatment for patients at an early stage of the disease.

## 9.4.2. Inria International Partners

#### 9.4.2.1. Informal International Partners

- Collaboration with the Department of Computer Science, University of Verona: Emmanuel Caruyer visited the group of Gloria Menegaz and Alessandro Daducci in the context of the 2017 School on Brain Connectomics (http://brainconnectomics.org/).
- Collaboration with Neuropoly, Polytechnique Montreal: Haykel Snoussi is visiting the group of Julien Cohen-Adad and received an Inria-MITACS fellowship for a 3 months period (Nov. 2017-Jan. 2018). He will be working on the processing of diffusion-weighted images of multiple sclerosis patients' spinal cord in the context of the EMISEP project.
- Collaboration with Department of Mathematics and Statistics at the Politechnico di Milano, Italy (Simone Vantini, Aymeric Stamm): Lorenzo Rota did visit the team between Oct. 2016 to March 2017 for his Tesi (Master degree) on "Application of shape analysis and functional data analysis tools on fiber bundles analysis".

## 9.5. International Research Visitors

## 9.5.1. Visits of International Scientists

• Simon Warfield and Benoit Scherrer, Harvard University, visited the VisAGeS team for the annual seminar on Jun. 2017.

## 9.5.2. Visits to International Teams

- Sudhanya Chatterjee visited the Computational Radiology Lab, the Boston Children's Hospital, at Harvard University in Nov. 2017. This stay was funded by the international program of University of Rennes 1. Christian Barillot and Olivier Commowick visited the same lab for a 3 days workshop in the context of the Associate Team.
- Haykel Snoussi visited the NeuroPoly Lab for 3 months from Nov. 2017. This stay was funded by the international program of University of Rennes 1.

## **10. Dissemination**

## **10.1.** Promoting Scientific Activities

## 10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

- Christian Barillot is member of the Board of Directors of IPMI conference series (Information Processing in Medical Imaging)
- Gilles Edan did organized the 25th anniversary workshop of the "Société Française de Neurologie" in Rennes on Nov. 30th

## 10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

- Christian Barillot was area chair of SPIE-MI and ISBI 2017
- Emmanuel Caruyer was Program Committee member of the CDMRI MICCAI workshop.

#### 10.1.2.2. Reviewer

• ISBI (Emmanuel Caruyer, Olivier Commowick), ISMRM (Élise Bannier), MICCAI (Emmanuel Caruyer, Olivier Commowick, Christian Barillot), IPMI (Olivier Commowick, Christian Barillot).

## 10.1.3. Journal

#### 10.1.3.1. Member of the editorial boards

- Christian Barillot is member of Editorial Boards of Medical Image Analysis, Current Medical Imaging Reviews, ISRN Signal Processing
- Christian Barillot is Editor-in-Chief of Frontiers in ICT: Computer Image Analysis.

#### 10.1.3.2. Reviewing Activities

• Am J Neuroradiol (Élise Bannier), Comput Biol Med (Christian Barillot), Comput Meth Prog Bio (Christian Barillot), Front Neurosc (Pierre Maurel), Hum Brain Mapp (Emmanuel Caruyer), JMRI (Isabelle Corouge), IEEE TMI (Pierre Maurel, Olivier Commowick), Med Image Anal (Olivier Commowick), Med Phys (Christian Barillot) Neuroimage (Christian Barillot, Isabelle Corouge, Emmanuel Caruyer, Olivier Commowick), Pattern Recog Lett (Christian Barillot).

## 10.1.4. Leadership within the Scientific Community

- Gilles Edan was elected Fellow of the European Academy of Neurologie. Member of the EAN teaching committee in 2015
- Christian Barillot is member of the Scientific Council of the INS2I<sup>0</sup> Institute of CNRS since 2011 and is Chairman of the Board since 2015
- Christian Barillot is member of the C3N committee (CNRS)
- Christian Barillot is member of the scientific board of "GIS France Grilles"
- Christian Barillot is member of the scientific board of the Neuroscience and psychiatry institute of AVIESAN

## 10.1.5. Scientific Expertise

- Christian Barillot provided an expertise for the Royal Netherlands Academy of Arts and Sciences (KNAW)
- Christian Barillot provided expertise for the Austrian Science Fund (FWF), the NSERC / CRSNG Canada, the Wellcome Trust (UK)
- Christian Barillot provided an expertise for the KU Leuven, the University of British Columbia, Vancouver, Canada
- Christian Barillot provided expertise for the Idex Université Grenoble Alpes
- Emmanuel Caruyer provided expertise for the Inria Associate Team program.

## 10.1.6. Invited Talks

- Emmanuel Caruyer gave an invited lecture, "Validating tractography pipelines: the help of simulated phantoms" during the 2017 School on Brain Connectomics, Oct. 9th-13th at the University of Verona, Italy.
- Emmanuel Caruyer gave an invited lecture, "Validating tractography pipelines with the help of simulated phantoms" during the Computational Brain Connectivity Mapping winter school workshop, Nov. 20th-24th in Juan-les-Pins, France.
- Yao Chi and Isabelle Corouge "Infrastructures pour le traitement et la gestion de données d'imagerie biologique et médicale ", Gen2bio National congress, Nantes, March 2017

<sup>&</sup>lt;sup>0</sup>http://csins2i.irisa.fr

## **10.2. Teaching - Supervision - Juries**

## 10.2.1. Teaching

- Emmanuel Caruyer, Introduction to algorithms (33h), L3 SIF, ENS Rennes, France
- Antoine Legouhy, Introduction to statistical modeling and R language (18h), L2 Biology, Univ. Rennes, France
- Antoine Legouhy, Introduction to numerical analysis and Python language (40h), L1 Biology, Univ. Rennes, France
- Corentin Vallée, Introduction to Machine learning (12h), M1 Fundamental and Applied Microbiology, Univ. Rennes, France
- Corentin Vallée, Machine learning (24h), M1 Molecular and Cellular biology, Univ. Rennes, France
- Corentin Vallée, Machine learning (24h), M1 Bioinformatics, Univ. Rennes, France
- Christian Barillot, Élise Bannier, Emmanuel Caruyer, Olivier Commowick, Isabelle Corouge, Jean-Yves Gauvrit, Master SIBM, University of Angers-Brest-Rennes (26h)
- Christian Barillot is responsible for one semester
- Jean-Yves Gauvrit is the coordinator for the Master
- École Supérieure d'Ingénieur de Rennes (ESIR): Pierre Maurel, General image processing (60h), Algorithmics and complexity (60h), Medical imaging (60h)
- ENS Rennes: Pierre Maurel, Introduction to image processing (24h)
- ISTIC Université of Rennes 1: Emmanuel Caruyer, Software Engineering (12h)

## 10.2.2. Supervision

- PhD in progress: Sudhanya Chatterjee, "Image-based Tissue Compartment Characterization of Neural Circuits with in-vivo MRI", Inria, from Nov 2015, Christian Barillot, Olivier Commowick, Jean-Christophe Ferré, Simon Warfield.
- PhD in progress: Cédric Meurée, "Quantitative Analysis Of Arterial Spin Labeling MRI For Robust Parametric Information Of Perfusion Maps", Inria / Siemens, from Mar 2014, Christian Barillot, Pierre Maurel.
- PhD in progress : Antoine Legouhy, "Longitudinal brain altas creation, application to development sudies", CNRS, from Nov 2016, Christian Barillot, François Rousseau, Olivier Commowick.
- PhD in progress : Corentin Vallée, "Joint estimation of neuronal activation, resting-state and basal metabolism from Arterial Spin Labeling", Univ. Rennes, from Nov 2016, Christian Barillot, Isabelle Corouge, Pierre Maurel.
- PhD in progress: Haykel Snoussi, "Diffusion MRI detection of early occurring spine lesions in relapsing-remitting multiple sclerosis with late physical impairment", from Nov 2015, Christian Barillot, Gilles Edan, Emmanuel Caruyer
- PhD in progress: Pierre-Yves Jonin, "Relationships between context-free and context-rich memory: cognitive and neural subtrates", Inria/Inserm/CNRS from Oct 2014, Christian Barillot (cosupervisor)
- PhD in progress: Maia Proisy, "Perfusion in neonates and in pediatric diseases", Univ. Rennes 1/CHRU Rennes, from Oct 2014, Jean-Christophe Ferré (supervisor)
- PhD in progress: Anne Kerbrat, "Quantitative MR imaging in MS for Brain and Spine", Univ. Rennes 1/CHRU Rennes, from Oct 2014, Gilles Edan (supervisor)
- PhD in progress: Stephanie Leplaideur, "As part of an fMRI substudy of a national research study on balance disorder rehabilitation", Univ. Rennes 1/CHRU Rennes, from Oct 2017, Isabelle Bonan (supervisor)

## 10.2.3. Juries

- Olivier Commowick, PhD committee : Renaud Hedouin, Inria, Rennes; June 12th, 2017.
- Christian Barillot, PhD committee: Renaud Hedouin, Inria, Rennes; June 12th, 2017.
- Christian Barillot, PhD committee: Lorraine Perronnet, Inria, Rennes; September 7th, 2017;
- Christian Barillot, PhD committee: Chunfeng LIAN, Rouen, January 27, 2017
- Christian Barillot, PhD review: Mehdi Hadj-Hamou, Nice Sophia Antipolis, January 17, 2017
- Christian Barillot, PhD review: Marco Pizzolato, Nice Sophia Antipolis, March 31, 2017
- Christian Barillot, HDR review: Julien Lefevre, Marseille, March 30, 2017
- Christian Barillot, PhD review: Riccardo Pascuzzo, Mathematics Dept., Politechnico di Milano, Italy, Dec. 2017

## **10.3.** Popularization

## 10.3.1. My thesis in 180 seconds

- Lorraine Perronnet and Cédric Meurée participated in the 2017 edition of the French edition of the Three Minutes Thesis competition" :
  - Lorraine Perronnet : "Combinaison de l'ElectroEncéphaloGraphie et de l'Imagerie par Résonance Magnétique fonctionnelle pour la réducation du cerveau par Neurofeedback"
  - Cédric Meurée : "Amélioration de la résolution d'images d'IRM de perfusion"

## 10.3.2. La semaine du cerveau (The Brain Week), from 13/03/2017 to 19/03/2017, Rennes

- Antoine Legouhy et Corentin Vallée : "PhD students processing MRI acquisition"
- Christophe Paya : "Discover brain anatomy with 3D models"
- Isabelle Corouge : "Neurinfo: An MRI dedicated to research"
- Pierre Maurel : "Let's play to the brain quizz in 3D immersion (using oculus rift)"
- Emmanuel Caruyer : "Medinria: Computer science serving cerebral imaging"

## 10.3.3. Journées Française de Radiologie 2017, 13-16 Octobre, Paris

- Stand FLI-IAM (Michael Kain, Yao Chi, Aneta Morawin, Mathieu Simon, Ines Fahkfahk, Arnaud Touboulic, Christian Barillot)
- Journée du Réseau d'Entraide en IRM Multicentrique (REMI) co-organisée par Elise Bannier

## 10.3.4. 50 ans Inria, Paris

• Stand "MedInria" (Olivier Commowick, Pierre Maurel), Nov 7-8, 2017

## 10.3.5. Journée du Président de la Société Française de Neurologie, Rennes, 7 Décembre 2017

• Organised by Gilles Edan

## 11. Bibliography

## Major publications by the team in recent years

[1] C. BARILLOT, E. BANNIER, O. COMMOWICK, I. COROUGE, A. BAIRE, I. FACKFACK, J. GUILLAUMONT, Y. YAO, M. KAIN.Shanoir: Applying the Software as a Service Distribution Model to Manage Brain Imaging Research Repositories, in "Frontiers in information and communication technologies", October 2016 [DOI: 10.3389/FICT.2016.00025], http://www.hal.inserm.fr/inserm-01404864.

- [2] C. BARILLOT, G. EDAN, O. COMMOWICK. Imaging biomarkers in Multiple Sclerosis: from image analysis to population imaging, in "Medical Image Analysis", 2016 [DOI: 10.1016/J.MEDIA.2016.06.017], https:// hal.inria.fr/hal-01333583.
- [3] O. COMMOWICK, A. MAAROUF, J.-C. FERRÉ, J.-P. RANJEVA, G. EDAN, C. BARILLOT.Diffusion MRI abnormalities detection with orientation distribution functions: A multiple sclerosis longitudinal study, in "Medical Image Analysis", May 2015, vol. 22, n<sup>o</sup> 1, p. 114-123 [DOI: 10.1016/J.MEDIA.2015.02.005], http://www.hal.inserm.fr/inserm-01134107.
- [4] R. HEDOUIN, O. COMMOWICK, E. BANNIER, B. SCHERRER, M. TAQUET, S. K. WARFIELD, C. BARILLOT.Block-Matching Distortion Correction of Echo-Planar Images With Opposite Phase Encoding Directions, in "IEEE Transactions on Medical Imaging", 2017, vol. 36, n<sup>o</sup> 5, p. 1106 - 1115 [DOI: 10.1109/TMI.2016.2646920], https://hal.inria.fr/hal-01436561.
- [5] M. MANO, A. LÉCUYER, E. BANNIER, L. PERRONNET, S. NOORZADEH, C. BARILLOT. *How to Build a Hybrid Neurofeedback Platform Combining EEG and fMRI.*, in "Frontiers in Neuroscience", March 2017, vol. 11, 140 [DOI: 10.3389/FNINS.2017.00140], http://www.hal.inserm.fr/inserm-01576500.
- [6] C. MAUMET, P. MAUREL, J.-C. FERRÉ, C. BARILLOT. An a contrario approach for the detection of patientspecific brain perfusion abnormalities with arterial spin labelling, in "NeuroImage", July 2016, vol. 134 [DOI: 10.1016/J.NEUROIMAGE.2016.03.054], http://www.hal.inserm.fr/inserm-01291748.
- [7] H. RAOULT, E. BANNIER, P. MAUREL, C. NEYTON, J.-C. FERRÉ, P. SCHMITT, C. BARILLOT, J.-Y. GAUVRIT.*Hemodynamic Quantification in Brain Arteriovenous Malformations With Time-Resolved Spin-Labeled Magnetic Resonance Angiography*, in "Stroke", July 2014, vol. 45, n<sup>O</sup> 8, p. 2461-4 [DOI: 10.1161/STROKEAHA.114.006080], http://www.hal.inserm.fr/inserm-01080106.

## **Publications of the year**

## **Doctoral Dissertations and Habilitation Theses**

- [8] R. HEDOUIN. *Diffusion MRI processing for multi-compartment characterization of brain pathology*, Université rennes1, June 2017, https://hal-univ-rennes1.archives-ouvertes.fr/tel-01548337.
- [9] L. PERRONNET. Combining EEG and FMRI for Neurofeedback, University of Rennes I, September 2017, https://hal.inria.fr/tel-01598667.
- [10] L. PERRONNET. Combining electroencephalography and functional magnetic resonance imaging for neurofeedback, Université Rennes 1, September 2017, https://tel.archives-ouvertes.fr/tel-01661583.

#### **Articles in International Peer-Reviewed Journal**

- [11] J. BONENFANT, E. BAJEUX, V. DEBURGHGRAEVE, E. LE PAGE, G. EDAN, A. KERBRAT. Can we stop immunomodulatory treatments in secondary progressive multiple sclerosis?, in "European Journal of Neurology", 2017, vol. 24, n<sup>o</sup> 2, p. 237-244 [DOI: 10.1111/ENE.13181], https://hal-univ-rennes1.archivesouvertes.fr/hal-01484970.
- [12] A. CARASS, S. ROY, A. JOG, J. L. CUZZOCREO, E. MAGRATH, A. GHERMAN, J. BUTTON, J. NGUYEN, F. PRADOS, C. H. SUDRE, M. JORGE CARDOSO, N. CAWLEY, O. CICCARELLI, C. A. WHEELER-KINGSHOTT, S. OURSELIN, L. CATANESE, H. DESHPANDE, P. MAUREL, O. COMMOWICK, C. BARILLOT,

X. TOMAS-FERNANDEZ, S. K. WARFIELD, S. VAIDYA, A. CHUNDURU, R. MUTHUGANAPATHY, G. KRISHNAMURTHI, A. JESSON, T. ARBEL, O. MAIER, H. HANDELS, L. O. IHEME, D. UNAY, S. JAIN, D. M. SIMA, D. SMEETS, M. GHAFOORIAN, B. PLATEL, A. BIRENBAUM, H. GREENSPAN, P.-L. BAZIN, P. A. CALABRESI, C. M. CRAINICEANU, L. M. ELLINGSEN, D. S. REICH, J. L. PRINCE, D. L. PHAM.Longitudinal multiple sclerosis lesion segmentation: Resource and challenge, in "NeuroImage", January 2017, vol. 148, p. 77 - 102 [DOI: 10.1016/J.NEUROIMAGE.2016.12.064], http://www.hal.inserm. fr/inserm-01480156.

- [13] B. COMBÈS, A. KERBRAT, O. COMMOWICK, A. MAAROUF, E. BANNIER, J. C. FERRÉ, A. TOURBAH, J.-P. RANJEVA, C. BARILLOT, G. EDAN. USPIO-positive MS lesions are associated with greater tissue damage than gadolinium-positive-only lesions during 3-year follow-up, in "Multiple Sclerosis", August 2017 [DOI: 10.1177/1352458517736148], https://hal.inria.fr/hal-01647039.
- [14] G. EDAN. Journée Société française de neurologie au JNLF 2017 : « Excellence in Neurology 2016 » [French Society of Neurology award: Excellence in Neurology 2016], in "Revue Neurologique", March 2017, vol. 173, n<sup>o</sup> 2, p. S33-S34 [DOI: 10.1016/J.NEUROL.2017.02.011], https://hal-univ-rennes1.archives-ouvertes. fr/hal-01577009.
- [15] G. EDAN, A. KERBRAT. Inhibitor of Fatty Acid Amide Hydrolase Learning from Tragic Failures, in "New England Journal of Medicine", 2017, vol. 376, p. 393-394 [DOI: 10.1056/NEJMc1615417], https://haluniv-rennes1.archives-ouvertes.fr/hal-01502531.
- [16] O. FERNÁNDEZ, M. DELVECCHIO, G. EDAN, S. FREDRIKSON, G. GIONVANNONI, P. HARTUNG, E. HAVRDOVA, L. KAPPOS, C. POZZILLI, P. SOERENSEN, B. TACKENBERG, P. VERMERSCH, G. COMI.Survey of diagnostic and treatment practices for multiple sclerosis in Europe, in "European Journal of Neurology", March 2017, vol. 24, n<sup>o</sup> 3, p. 516 - 522 [DOI: 10.1111/ENE.13236], http://www.hal.inserm. fr/inserm-01576518.
- [17] K. C. FITZGERALD, K. L. MUNGER, H.-P. HARTUNG, M. S. FREEDMAN, X. MONTALBÁN, G. EDAN, E.-M. WICKLEIN, E.-W. RADUE, L. KAPPOS, C. POHL, A. ASCHERIO. Sodium intake and multiple sclerosis activity and progression in BENEFIT, in "Annals of Neurology", July 2017, vol. 82, n<sup>o</sup> 1, p. 20 - 29 [DOI: 10.1002/ANA.24965], http://www.hal.inserm.fr/inserm-01576520.
- [18] B. GORY, L. HUOT, R. RIVA, P. LABEYRIE, O. LEVRIER, A. LEBEDINSKY, H. BRUNEL, J.-Y. GAU-VRIT, R. BLANC, E. CHABERT, L. DEREX, E. EMERY, A. NICOLAS, H. DESAL, G. RODESCH, F. TURJMAN.One-year efficacy and safety of the Trufill DCS Orbit and Orbit Galaxy detachable coils in the endovascular treatment of intracranial aneurysms: Results from the TRULINE study, in "Interventional Neuroradiology", April 2017 [DOI: 10.1177/1591019917717576], https://hal.inria.fr/hal-01576317.
- [19] J. HASCOET, A. MANUNTA, C. BROCHARD, A. ARNAUD, M. DAMPHOUSSE, H. MÉNARD, J. KERDRAON, H. JOURNEL, I. BONAN, S. ODENT, B. FREMOND, L. SIPROUDHIS, X. GAMÉ, B. PEYRONNET. Outcomes of intra-detrusor injections of botulinum toxin in patients with spina bifida: A systematic review, in "Neurourology and Urodynamics", 2017, vol. 36, p. 557-564 [DOI : 10.1002/NAU.23025], https://hal-univ-rennes1. archives-ouvertes.fr/hal-01502530.
- [20] M. HATT, B. LAURENT, A. OUAHABI, H. FAYAD, S. TAN, L. LI, W. LU, V. JAOUEN, C. TAUBER, J. CZAKON, F. DRAPEJKOWSKI, W. DYRKA, S. CAMARASU-POP, F. CERVENANSKY, P. GIRARD, T. GLATARD, M. KAIN, Y. YAO, C. BARILLOT, A. KIROV, D. VISVIKIS. *The first MICCAI challenge on PET tumor segmentation*, in "Medical Image Analysis", 2017, p. 1-46 [DOI: 10.1016/J.MEDIA.2017.12.007], https://hal.archives-ouvertes.fr/hal-01659162.

- [21] R. HEDOUIN, O. COMMOWICK, E. BANNIER, B. SCHERRER, M. TAQUET, S. K. WARFIELD, C. BARILLOT.Block-Matching Distortion Correction of Echo-Planar Images With Opposite Phase Encoding Directions, in "IEEE Transactions on Medical Imaging", 2017, vol. 36, n<sup>o</sup> 5, p. 1106 - 1115 [DOI: 10.1109/TMI.2016.2646920], https://hal.inria.fr/hal-01436561.
- [22] H. HUGUES, M. DI MARCO, J. JANIAUD, Y. XUE, J. PIRES, H. KHADEMI, M. CUCHERAT, I. BONAN, G. GUEYFFIER, G. RODE. Efficiency of physical therapy on postural imbalance after stroke: study protocol for a systematic review and meta-analysis, in "BMJ Open", January 2017, vol. 7, n<sup>O</sup> 1 [DOI: 10.1136/BMJOPEN-2016-013348], http://www.hal.inserm.fr/inserm-01576495.
- [23] S. KALSON-RAY, G. EDAN, E. LERAY.An excessive risk of suicide may no longer be a reality for multiple sclerosis patients, in "Multiple Sclerosis", 2017, vol. 23, n<sup>o</sup> 6, p. 864-871 [DOI: 10.1177/1352458517699873], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01518889.
- [24] D. LE FOLL, H. RAOULT, J. FERRÉ, F. NAUDET, D. TRYSTRAM, J. GAUVRIT. The learning curve effect on embolization of cerebral dural arteriovenous fistula single-center experience in 48 consecutive patients, in "American Journal of Neuroradiology", February 2017, vol. 44, n<sup>o</sup> 1, p. 52 56 [DOI: 10.1016/J.NEURAD.2016.09.008], https://hal.inria.fr/hal-01557069.
- [25] E. LE PAGE, G. EDAN.Oral rather than intravenous corticosteroids should be used to treat MS relapses
  Yes, in "Multiple Sclerosis", 2017, vol. 23, n<sup>o</sup> 8, p. 1056–1058 [DOI: 10.1177/1352458517695893], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01559941.
- [26] K. H. MAIER-HEIN, P. F. NEHER, C. HOUDE, M.-A. CÔTÉ, E. GARYFALLIDIS, J. ZHONG, M. CHAMBER-LAND, F.-C. YEH, Y.-C. LIN, Q. JI, W. E. REDDICK, J. O. GLASS, D. QIXIANG CHEN, Y. FENG, C. GAO, Y. WU, J. MA, H. RENJIE, Q. LI, C.-F. WESTIN, S. DESLAURIERS-GAUTHIER, J. O. O. GONZÁLEZ, M. PAQUETTE, S. ST-JEAN, G. GIRARD, F. RHEAULT, J. SIDHU, C. M. W. TAX, F. GUO, H. Y. MESRI, S. DÁVID, M. FROELING, A. M. HEEMSKERK, A. LEEMANS, A. BORÉ, B. PINSARD, C. BEDETTI, M. DESROSIERS, B. BRAMBATI, J. DOYON, A. SARICA, R. VASTA, A. CERASA, A. QUATTRONE, J. YEATMAN, A. R. KHAN, W. HODGES, S. ALEXANDER, D. ROMASCANO, M. BARAKOVIC, A. AURÍA, O. ESTEBAN, A. LEMKADDEM, J.-P. THIRAN, H. E. CETINGUL, B. L. ODRY, B. MAILHÉ, M. S. NADAR, F. M. PIZZAGALLI, G. PRASAD, J. E. VILLALON-REINA, J. GALVIS, P. M. THOMPSON, F. D. S. REQUEJO, P. L. LAGUNA, L. M. LACERDA, R. BARRETT, F. DELL'ACQUA, M. CATANI, L. PETIT, E. CARUYER, A. DADUCCI, T. DYRBY, T. HOLLAND-LETZ, C. C. HILGETAG, B. STIELTJES, M. DESCOTEAUX.*The challenge of mapping the human connectome based on diffusion tractography*, in "Nature Communications", December 2017, vol. 8, n<sup>O</sup> 1 [DOI: 10.1038/s41467-017-01285-x], https://hal.inria.fr/hal-01631578.
- [27] M. MANO, A. LÉCUYER, E. BANNIER, L. PERRONNET, S. NOORZADEH, C. BARILLOT.*How to Build a Hybrid Neurofeedback Platform Combining EEG and fMRI*, in "Frontiers in Neuroscience", March 2017, vol. 11, 140 [DOI: 10.3389/FNINS.2017.00140], http://www.hal.inserm.fr/inserm-01576500.
- [28] C. MERCK, I. COROUGE, P.-Y. JONIN, B. DESGRANGES, J.-Y. GAUVRIT, S. BELLIARD.What semantic dementia teaches us about the functional organization of the left posterior fusiform gyrus, in "Neuropsychologia", September 2017, vol. 106, p. 159-168 [DOI : 10.1016/J.NEUROPSYCHOLOGIA.2017.09.023], http://www.hal.inserm.fr/inserm-01613998.
- [29] A. PAISANT, A. BOULIC, E. BARDOU-JACQUET, E. BANNIER, G. D'ASSIGNIES, F. LAINÉ, B. TURLIN, Y. GANDON. Assessment of liver iron overload by 3 T MRI, in "Abdominal Radiology", February 2017, vol. 42, nº 6, p. 1713-1720 [DOI: 10.1007/s00261-017-1077-8], http://www.hal.inserm.fr/inserm-01477775.

- [30] R. PASCOT, A. DAOUDAL, A. CARDON, G. GODET, A. LUCAS, E. CLOCHARD, J.-Y. GAUVRIT, Y. LE TEURNIER, A. KALADJI. Evaluation by Magnetic Resonance Imaging of Silent Brain Infarcts in Preoperative and Postoperative Asymptomatic Carotid Surgery, in "Annals of Vascular Surgery", August 2017, vol. 43, p. 258-264 [DOI: 10.1016/J.AVSG.2016.10.067], http://www.hal.inserm.fr/inserm-01576504.
- [31] L. PERRONNET, A. LÉCUYER, M. MANO, E. BANNIER, F. LOTTE, M. CLERC, C. BARILLOT. Unimodal Versus Bimodal EEG-fMRI Neurofeedback of a Motor Imagery Task, in "Frontiers in Human Neuroscience", April 2017, vol. 11 [DOI : 10.3389/FNHUM.2017.00193], https://hal.inria.fr/hal-01519755.
- [32] M. PROISY, L. RIFFAUD, K. CHOUKLATI, C. TRÉGUIER, B. BRUNEAU. Ultrasonography for the diagnosis of craniosynostosis, in "European Journal of Radiology", May 2017, vol. 90, p. 250–255 [DOI: 10.1016/J.EJRAD.2017.03.006], https://hal.archives-ouvertes.fr/hal-01580574.

#### **International Conferences with Proceedings**

- [33] E. BANNIER, B. COMBES, L. MONTEAU, C. BARILLOT, G. EDAN, A. KERBRAT, J.-C. FERRE. *Evaluating the reliability of cervical spine magnetization transfer imaging in the context of multicenter longitudinal studies*, in "ESMRMB", Barcelona, Spain, October 2017, https://hal.inria.fr/hal-01578473.
- [34] O. COMMOWICK, R. HEDOUIN, E. CARUYER, C. BARILLOT.L2 Similarity Metrics for Diffusion Multi-Compartment Model Images Registration, in "20th International Conference on Medical Image Computing and Computer Assisted Intervention - MICCAI", Québec, Canada, LNCS, September 2017, vol. 10433, p. 257-265 [DOI: 10.1007/978-3-319-66182-7\_30], http://www.hal.inserm.fr/inserm-01556476.
- [35] M. MANO, E. BANNIER, L. PERRONNET, A. LÉCUYER, C. BARILLOT. Hybrid EEG and fMRI platform for multi-modal neurofeedback, in "International Society of Magnetic Resonance in Medicine", Honolulu, United States, Proc. Intl. Soc. Mag. Reson. Med. 25th, ISMRM, April 2017, 4550, http://www.hal.inserm.fr/inserm-01577442.
- [36] C. MEURÉE, P. MAUREL, E. BANNIER, C. BARILLOT. Patch-based super-resolution for arterial spin labeling MRI, in "ISMRM 25th Annual Meeting & Exhibition", Honolulu, United States, April 2017, http://www.hal. inserm.fr/inserm-01558183.
- [37] S. NOORZADEH, P. MAUREL, T. OBERLIN, R. GRIBONVAL, C. BARILLOT. Multi-modal EEG and fMRI Source Estimation Using Sparse Constraints, in "MICCAI 2017 - 20th International Conference on Medical Image Computing and Computer Assisted Intervention", Quebec, Canada, September 2017 [DOI: 10.1007/978-3-319-66182-7\_51], https://hal.inria.fr/hal-01586495.
- [38] H. SNOUSSI, E. CARUYER, O. COMMOWICK, E. BANNIER, C. BARILLOT. Comparison of inhomogeneity distortion correction methods in diffusion MRI of the spinal cord, in "ESMRMB - 34th Annual Scientific Meeting European Society for Magnetic Resonance in Medecine and Biology", Barcelone, Spain, October 2017, Scientific Session, http://www.hal.inserm.fr/inserm-01579134.

#### **Conferences without Proceedings**

[39] A.-C. BINTER, E. BANNIER, S. CORDIER, J.-F. VIEL, C. CHEVRIER, F. PELÉ. Prenatal solvent exposure and motor inhibition function at the age of 10 – 12 in the PELAGIE mother-child cohort, in "International Society of Environmental Epidemiology", Sydney, Australia, September 2017, http://www.hal.inserm.fr/ inserm-01578507.

- [40] A.-C. BINTER, E. BANNIER, D. SAINT-AMOUR, G. SIMON, J.-C. FERRÉ, C. BARILLOT, J.-F. VIEL, S. CORDIER, C. CHEVRIER, F. PELÉ. Prenatal solvent exposure association with motor inhibition functional MRI at the age of 10-12 in the PELAGIE mother-child cohort, in "International Society of Environmental Epidemiology", Sydney, Australia, September 2017, http://www.hal.inserm.fr/inserm-01578506.
- [41] S. CHATTERJEE, O. COMMOWICK, S. K. WARFIELD, C. BARILLOT. Gaining Insights Into Multiple Sclerosis Lesion Characteristics from Brain Tissue Microstructure Information: A Multi-Compartment T2 Relaxometry Approach, in "ISMRM 25TH ANNUAL MEETING & EXHIBITION", Honolulu, Hawaii, USA, United States, April 2017, http://www.hal.inserm.fr/inserm-01543061.
- [42] S. CHATTERJEE, O. COMMOWICK, S. K. WARFIELD, C. BARILLOT. Multi-Compartment T2 Relaxometry Model Using Gamma Distribution Representations: A Framework for Quantitative Estimation of Brain Tissue Microstructures, in "ISMRM 25TH ANNUAL MEETING & EXHIBITION", Honolulu, Hawaii, USA, United States, April 2017, http://www.hal.inserm.fr/inserm-01543073.
- [43] A. KERBRAT, B. COMBÈS, L. MONTEAU, J. FERRÉ, J. MARANZANO, V. CALLOT, P. LABAUGE, X. AYRIGNAC, C. CARRA DALLIÈRE, N. MENJOT DE CHAMPFLEUR, J.-P. PELLETIER, A. MAAROUF, D. BRASSAT, J. DE SEZE, N. COLLONGUES, F. DURAND DUBIEF, C. BARILLOT, G. EDAN, E. BAN-NIER. Presence of focal and diffuse cervical cord tissue damage in early relapsing-remitting MS: A magnetization transfer study, in "ECTRIMS", Paris, France, October 2017, https://hal.inria.fr/hal-01579486.