Activity Report Rennes - Bretagne Atlantique 2019

Edition: 2020-03-23

List of Inria's Research Teams

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

Project-Team CAIRN

Creation of the Project-Team: 2009 January 01

Keywords:

Computer Science and Digital Science:

A1.1. - Architectures
A1.1.1. - Multicore, Manycore
A1.1.2. - Hardware accelerators (GPGPU, FPGA, etc.)
A1.1.8. - 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
A2.2.6. - GPGPU, FPGA...
A2.2.7. - 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. Team, Visitors, External Collaborators

Research Scientists

François Charot [Inria, Researcher]
Silviu-Ioan Filip [Inria, Researcher]
Tomofumi Yuki [Inria, Researcher]

Faculty Members

Olivier Sentieys [Team leader, Professor, Univ. Rennes, Inria Chair, HDR]
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, Rennes, HDR]

Technical Staff

Mickaël Le Gentil [Research Engineer (half time), Univ. Rennes, Lannion]
Justine Bonnot [Inria, Engineer, from Nov. 2019]
Pierre Hallé [Inria, Engineer, Lannion]
Simon Rokicki [Inria, Engineer, Rennes]

PhD Students

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. Rennes, granted by USTH, Rennes, from May 2017]
Thibaut Marty [Univ. Rennes, granted by H2020 ARGO and Brittany Region, Rennes, from Sep. 2017]
Petr Dobias [Univ. Rennes, MENRT grant, Lannion, from Oct. 2017]
Van-Phu Ha [Inria, granted by ANR Artefact, Rennes, from Nov. 2017]
Romain Mercier [Inria, granted by DGA and Inria, Lannion, from Oct. 2018]
Minyu Cui [Univ. Rennes, granted by CSC-ENS, Rennes, from Sep. 2018]
Jaechul Lee [Univ. Rennes, granted by Brittany Region/LTC, Lannion, from Dec. 2018]
Davide Pala [Inria, granted by IPL ZEP, Rennes, from Jan. 2018]
Joseph Paturel [Inria, granted by RAPID FLODAM, Rennes, from Sep. 2018]
Thibault Allenet [Univ. Rennes, granted by CEA LIST, Saclay, from March 2019]
Corentin Ferry [Univ. Rennes, granted by ENS, Rennes, from Sep. 2019]
Adrien Gaonac'h [Univ. Rennes, granted by CEA LIST, Saclay, from Oct. 2019]
Genevieve Ndour [Univ. Rennes, granted by CEA Leti, Grenoble, from May 2016]
Audrey Lucas [CNRS, granted by DGA-PEC, Lannion, until Jan. 2019]

Post-Doctoral Fellows

Mansureh Shahraki Moghaddam [Inria, Rennes, until May 2019]
Lei Mo [Inria, Rennes, until Apr 2019]

Visiting Scientists

Bernard Goossens [Professor, Univ. Perpignan, July 2019]
Sharad Sinha [Professor, IIT Goa, India, July 2019]

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) [51] and then by Coarse-Grain Reconfigurable Arrays (CGRA) [54], [66] [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 [57], 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 [61], 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 [55]. 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 have been advocating 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. Time has confirmed the importance of heterogeneous manycore architectures, which are prevalent today.

> 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 to tackle these challenges, leveraging the background and past experience of the team members. 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.

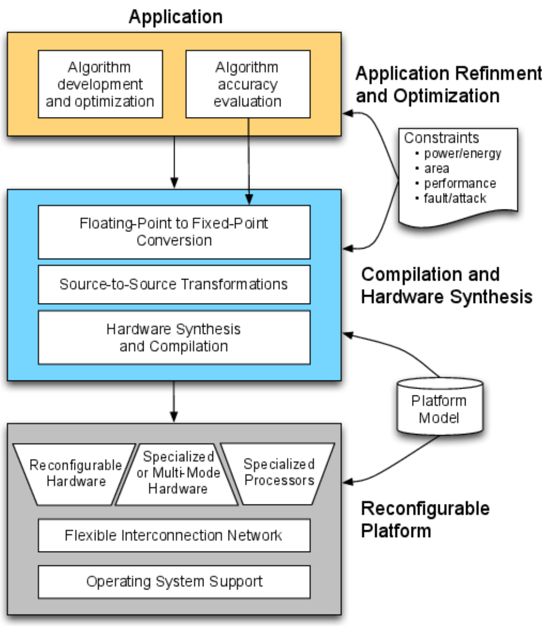


Figure 1. CAIRN*'s general design flow and related research themes*

Figure 1 shows the global design flow we propose to develop. This flow is organized in levels corresponding 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 [69]. They can also be reconfigured dynamically. 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 related work has been 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 [53] or implementing an online filter to relocate the tasks [63]. 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. Coarse-grained reconfigurable architectures (CGRA) provide operator-level configurable functional blocks and word-level datapaths [70], [58], [68]. 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[9], Adres [66] or polymorphous computing fabrics[11]. These works have led to commercial products such as the PACT/XPP [52] 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 general-purpose 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 low-power 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 [56], [71], [67], [65], [64]. 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 [62]. Here again, existing tools are not well suited, as they fail at providing a 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) [5], ASIP optimizing compilers [12] 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 [6]. 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 [4].
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 floating-point 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 [7] 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 [59]. 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 [60]. 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 [10].

4. Application Domains

4.1. Panorama

keywords: Wireless (Body) Sensor Networks, High-Rate Optical Communications, Wireless Communications, Applied Cryptography, Machine Learning, Deep Learning, Image ans Signal Processing.
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, data mining, and multimedia processing.

5. New Software and Platforms

5.1. 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

5.2. 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 in 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 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 optimized 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 constructs an analytic accuracy model of the program, which means it can explore more solutions and thereby produce a much more efficient code.

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

5.3. SmartSense

KEYWORDS: Wireless Sensor Networks - Smart building - Non-Intrusive Appliance Load Monitoring

FUNCTIONAL DESCRIPTION: To measure energy consumption by equipment in a building, NILM techniques (Non-Intrusive Appliance Load Monitoring) are based on observation of overall variations in electrical voltage. This avoids having to deploy watt-meters on every device and thus reduces the cost. SmartSense goes a step further to improve on these techniques by combining sensors (light, temperature, electromagnetic wave, vibration and sound sensors, etc.) to provide additional information on the activity of equipment and people. Low-cost sensors can be energy-autonomous too.

- Contact: Olivier Sentieys

5.4. Platforms

5.4.1. Zyggie: a Wireless Body Sensor Network Platform

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 (see Figure 2) includes the following features: a 32-bit micro-controller to manage a custom MAC layer and process 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: https://bowi.cominlabs.u-bretagneloire.fr/zyggie-wbsn-platform

Figure 2. CAIRN's Zyggie platform for WBSN

5.4.2. E-methodHW: an automatic tool for the evaluation of polynomial and rational function approximations

KEYWORDS: function approximation, FPGA hardware implementation generator

SCIENTIFIC DESCRIPTION: E-methodHW is an open source C/C++ prototype tool written to exemplify what kind of numerical function approximations can be developed using a digit recurrence evaluation scheme for polynomials and rational functions.

FUNCTIONAL DESCRIPTION: E-methodHW provides a complete design flow from choice of mathematical function operator up to optimised VHDL code that can be readily deployed on an FPGA. The use of the E-method allows the user great flexibility if targeting high throughput applications.

- Participants: Silviu-Ioan Filip, Matei Istoan
- Partners: Université de Rennes 1, Imperial College London
- Contact: Silviu-Ioan Filip
- URL: https://github.com/sfilip/emethod

5.4.3. Firopt: a tool for the simultaneous design of digital FIR filters along with the dedicated hardware model

KEYWORDS: FIR filter design, multiplierless hardware implementation generator

SCIENTIFIC DESCRIPTION: the firopt tool is an open source C++ prototype that produces Finite Impulse Response (FIR) filters that have minimal cost in terms of digital adders needed to implement them. This project aims at fusing the filter design problem from a frequency domain specification with the design of the dedicated hardware architecture. The optimality of the results is ensured by solving appropriate mixed integer linear programming (MILP) models developed for the project. It produces results that are generally more efficient than those of other methods found in the literature or from commercial tools (such as MATLAB).

- Participants: Silviu-Ioan Filip, Martin Kumm, Anastasia Volkova
- Partners: Université de Rennes 1, Université de Nantes, Fulda University of Applied Sciences
- Contact: Silviu-Ioan Filip
- URL: https://gitlab.com/filteropt/firopt

5.4.4. Hybrid-DBT

KEYWORDS: Dynamic Binary Translation, hardware acceleration, VLIW processor, RISC-V

SCIENTIFIC DESCRIPTION: Hybrid-DBT is a hardware/software Dynamic Binary Translation (DBT) framework capable of translating RISC-V binaries into VLIW binaries. Since the DBT overhead has to be as small as possible, our implementation takes advantage of hardware acceleration for performance critical stages (binary translation, dependency analysis and instruction scheduling) of the flow. Thanks to hardware acceleration, our implementation is two orders of magnitude faster than a pure software implementation and enable an overall performance improvements by 23% on average, compared to a native RISC-V execution.

- Participants: Simon Rokicki, Steven Derrien
- Partners: Université de Rennes 1
- URL: https://github.com/srokicki/HybridDBT

5.4.5. Comet

KEYWORDS: Processor core, RISC-V instruction-set architecture

SCIENTIFIC DESCRIPTION: Comet is a RISC-V pipelined processor with data/instruction caches, fully developed using High-Level Synthesis. The behavior of the core is defined in a small C code which is then fed into a HLS tool to generate the RTL representation. Thanks to this design flow, the C description can be used as a fast and cycle-accurate simulator, which behaves exactly like the final hardware. Moreover, modifications in the core can be done easily at the C level.

- Participants: Simon Rokicki, Steven Derrien, Olivier Sentieys, Davide Pala, Joseph Paturel
- Partners: Université de Rennes 1
- URL: https://gitlab.inria.fr/srokicki/Comet

5.4.6. TypEx

KEYWORDS: Embedded systems, Fixed-point arithmetic, Floating-point, Low power consumption, Energy efficiency, FPGA, ASIC, Accuracy optimization, Automatic floating-point to fixed-point conversion

SCIENTIFIC DESCRIPTION: TypEx is a tool designed to automatically determine custom number representations and word-lengths (i.e., bit-width) for FPGAs and ASIC designs at the C source level. The main goal of TypEx is to explore the design space spanned by possible number formats in the context of High-Level Synthesis. TypEx takes a C code written using floating-point datatypes specifying the application to be explored. The tool also takes as inputs a cost model as well as some user constraints and generates a C code where the floating-point datatypes are replaced by the wordlengths found after exploration. The best set of word-lengths is the one found by the tool that respects the given accuracy constraint and that minimizes a parametrized cost function. Figure 3 presents an overview of the TypEx design flow.

- Participants: Olivier Sentieys, Tomofumi Yuki, Van-Phu Ha
- Partners: Université de Rennes 1
- URL: https://gitlab.inria.fr/gecos/gecos-float2fix

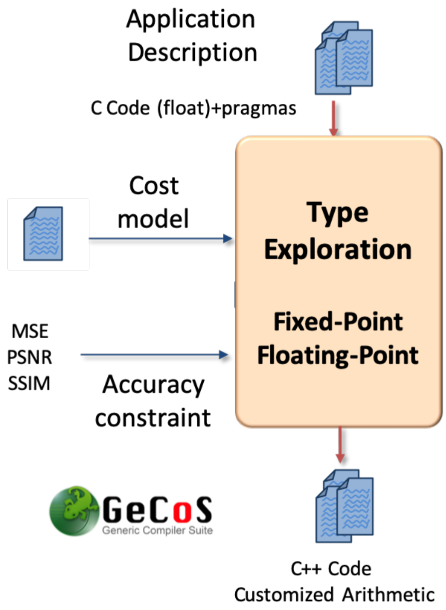


Figure 3. TypEx: a tool for type exploration and automatic floating-point to fixed-point conversion

6. New Results

6.1. Reconfigurable Architecture and Hardware Accelerator Design

6.1.1. Algorithmic Fault Tolerance for Timing Speculative Hardware

Participants: Thibaut Marty, Tomofumi Yuki, Steven Derrien.

We have been working on timing speculation, also known as overclocking, to increase the computational throughput of accelerators. However, aggressive overclocking introduces timing errors, which may corrupt the outputs to unacceptable levels. It is extremely challenging to ensure that no timing errors occur, since the probability of such errors happening depends on many factors including the temperature and process variation. Thus, aggressive timing speculation must be coupled with a mechanism to verify that the outputs are correctly computed. Our previous result demonstrated that the use of inexpensive checks based on algebraic properties of the computation can drastically reduce the cost of verifying that overclocking did not produce incorrect outputs. This has allowed the accelerator to significantly boost its throughput with little area overhead.

One weakness coming from the use of algebraic properties is that the inexpensive check is not strictly compatible with floating-point arithmetic that is not associative. This was not an issue with our previous work that targeted convolutional neural networks, which typically use fixed-point (integer) arithmetic. Our on-going work aims to extend our approach to floating-point arithmetic by using extended precision to store intermediate results, known as Kulisch accumulators. At first glance, use of extended precision that covers the full exponent range of floating-point may look costly. However, the design space of FPGAs is complex with many different trade-offs, making the optimal design highly context dependent. Our preliminary results indicate that the use of extended precision may not be any more costly than implementing the computation in floating point.

6.1.2. Adaptive Dynamic Compilation for Low-Power Embedded Systems

Participants: Steven Derrien, Simon Rokicki.

Previous works on Hybrid-DBT have demonstrated that using Dynamic Binary Translation, combined with low-power in-order architecture, enables an energy-efficient execution of compute-intensive kernels. In [33], we address one of the main performance limitations of Hybrid-DBT: the lack of speculative execution. We study how it is possible to use memory dependency speculation during the DBT process. Our approach enables fine-grained speculation optimizations thanks to a combination of hardware and software mechanisms. Our results show that our approach leads to a geo-mean speed-up of 10% at the price of a 7% area overhead. In [49], we summarize the current state of the Hybrid-DBT project and display our last results about the performance and the energy efficiency of the system. The experimental results presented here show that, for compute-intensive benchmarks, Hybrid-DBT can deliver the same performance level than a 3-issue OoO core, while consuming three times less energy. Finally, in [34], we investigate security issues caused by the use of speculation in DBT-based systems. We demonstrate that, even if those systems use in-order micro-architectures, the DBT layer optimizes binaries and speculates on the outcome of some branches, leading to security issues similar to the Spectre vulnerability. We demonstrate that both the NVidia Denver architecture and the Hybrid-DBT platform are subject to such vulnerability. However, we also demonstrate that those systems can easily be patched, as the DBT is done in software and has fine-grained control over the optimization process.

6.1.3. What You Simulate Is What You Synthesize: Designing a Processor Core from C++ Specifications

Participants: Simon Rokicki, Davide Pala, Joseph Paturel, Olivier Sentieys.

Designing the hardware of a processor core as well as its verification flow from a single high-level specification would provide great advantages in terms of productivity and maintainability. In [32] (a preliminary version also in [42]), we highlight the gain of starting from a unique high-level synthesis and simulation C++ model to design a processor core implementing the RISC-V Instruction Set Architecture (ISA). The specification code is used to generate both the hardware target design through High-Level Synthesis as well as a fast and cycle-accurate bit-accurate simulator of the latter through software compilation. The object oriented nature of C++ greatly improves the readability and flexibility of the design description compared to classical HDL-based implementations. Therefore, the processor model can easily be modified, expanded and verified using standard software development methodologies. The main challenge is to deal with C++ based synthesizable specifications of core and uncore components, cache memory hierarchy, and synchronization. In particular, the research question is how to specify such parallel computing pipelines with high-level synthesis technology and to demonstrate that there is a potential high gain in design time without jeopardizing performance and cost. Our experiments demonstrate that the core frequency and area of the generated hardware are comparable to existing RTL implementations.

6.1.4. Accelerating Itemset Sampling on FPGA

Participants: Mael Gueguen, Olivier Sentieys.

Finding recurrent patterns within a data stream is important for fields as diverse as cybersecurity or e-commerce. This requires to use pattern mining techniques. However, pattern mining suffers from two issues. The first one, known as "pattern explosion", comes from the large combinatorial space explored and is the result of too many patterns outputted to be analyzed. Recent techniques called output space sampling solve this problem by outputting only a sampled set of all the results, with a target size provided by the user. The second issue is that most algorithms are designed to operate on static datasets or low throughput streams. In [24], we propose a contribution to tackle both issues, by designing an FPGA accelerator for pattern mining with output space sampling. We show that our accelerator can outperform a state-of-the-art implementation on a server class CPU using a modest FPGA product. This work is done in collaboration with A. Termier from the Lacodam team at Inria.

6.1.5. Hardware Accelerated Simulation of Heterogeneous Platforms

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

When considering designing heterogeneous multicore 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 multicore architectures. These limitations motivate the use of hardware to accelerate the simulation of heterogeneous multicore, and in particular of FPGA components. We study an approach for designing such systems based on performance models through combining accelerator and processor core models. These models are implemented in the HAsim/LEAP infrastructure. In [22], we propose a methodology for building performance models of accelerators and describe the defined design flow.

6.1.6. Fault-Tolerant Scheduling onto Multicore embedded Systems

Participants: Emmanuel Casseau, Minyu Cui, Petr Dobias, Lei Mo, Angeliki Kritikakou.

Demand on multiprocessor 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. Temporal and/or spatial redundancy is currently used to tackle this issue. Actually, multiprocessor 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 map and schedule tasks onto homogeneous faulty processors.

We consider two approaches. The first approach deals with task mapping onto processors at compile time. Our goal is to guarantee both reliability and hard real-time constraints with low-energy consumption. Task duplication is assessed and duplication is performed if expected reliability of a task is not met. This work concurrently decides duplication of tasks, the task execution frequency and task allocation to minimize the energy consumption of a multicore platform with Dynamic Voltage and Frequency Scaling (DVFS) capabilities. The problem is initially formulated as Integer Non-Linear Programming and equivalently transformed to a Mixed Integer Linear Programming problem to be optimally solved. The proposed approach provides a good trade-off between energy consumption and reliability. The second approach deals with mapping and scheduling tasks at runtime. The application context is CubeSats. CubeSats operate in harsh space environment and they are exposed to charged particles and radiations, which cause transient faults. To make CubeSats fault tolerant, we propose to take advantage of their multicore architecture. We propose two online algorithms, which schedule all tasks on board of a CubeSat, detect faults and take appropriate measures (based on task replication) in order to deliver correct results. The first algorithm considers all tasks as aperiodic tasks and the second one treats them as aperiodic or periodic tasks. Their performances vary, particularly when the number of processors is low, and a choice is subject to a trade-off between the rejection rate and the energy consumption. This work is done in collaboration with Oliver Sinnen, PARC Lab., the University of Auckland.

6.1.7. Run-Time Management on Multicore Platforms

Participant: Angeliki Kritikakou.

In time-critical systems, run-time adaptation is required to improve the performance of time-triggered execution, derived based on Worst-Case Execution Time (WCET) of tasks. By improving performance, the systems can provide higher Quality-of-Service, in safety-critical systems, or execute other best-effort applications, in mixed-critical systems. To achieve this goal, we propose a parallel interference-sensitive run-time adaptation mechanism that enables a fine-grained synchronisation among cores [37]. Since the run-time adaptation of offline solutions can potentially violate the timing guarantees, we present the Response-Time Analysis (RTA) of the proposed mechanism showing that the system execution is free of timing-anomalies. The RTA takes into account the timing behavior of the proposed mechanism and its associated WCET. To support our contribution, we evaluate the behavior and the scalability of the proposed approach for different application types and execution configurations on the 8-core Texas Instruments TMS320C6678 platform. The obtained results show significant performance improvement compared to state-of-the-art centralized approaches.

6.1.8. Energy Constrained and Real-Time Scheduling and Assignment on Multicores

Participants: Olivier Sentieys, Angeliki Kritikakou, Lei Mo.

Asymmetric Multicore Processors (AMP) are a very promising architecture to deal efficiently with the wide diversity of applications. In real-time application domains, in-time approximated results are preferred to accurate – but too late – results. In [28], we propose a deployment approach that exploits the heterogeneity provided by AMP architectures and the approximation tolerance provided by the applications, so as to increase as much as possible the quality of the results under given energy and timing constraints. Initially, an optimal approach is proposed based on the problem linearization and decomposition. Then, a heuristic approach is developed based on iteration relaxation of the optimal version. The obtained results show 16.3% reduction in the computation time for the optimal approach compared to conventional optimal approaches. The proposed heuristic approach is about 100 times faster at the cost of a 29.8% QoS degradation in comparison with the optimal solution.

6.1.9. Real-Time Energy-Constrained Scheduling in Wireless Sensor and Actuator Networks

Participants: Angeliki Kritikakou, Lei Mo.

Cyber-Physical Systems (CPS), as a particular case of distributed systems, raise new challenges, because of the heterogeneity and other properties traditionally associated with Wireless Sensor and Actuator Networks (WSAN), including shared sensing, acting and real-time computing. In CPS, mobile actuators can enhance system's flexibility and scalability, but at the same time incur complex couplings in the scheduling and controlling of the actuators. In [19], we propose a novel event-driven method aiming at satisfying a required level of control accuracy and saving energy consumption of the actuators, while guaranteeing a bounded action delay. We formulate a joint-design problem of both actuator scheduling and output control. To solve this problem, we propose a two-step optimization method. In the first step, the problem of actuator scheduling and action time allocation is decomposed into two subproblems. They are solved iteratively by utilizing the solution of one in the other. The convergence of this iterative algorithm is proved. In the second step, an on-line method is proposed to estimate the error and adjust the outputs of the actuators accordingly. Through simulations and experiments,we demonstrate the effectiveness of the proposed method. In addition, many of the real-time tasks of CPS can be executed in an imprecise way. Such systems accept an approximate result as long as the baseline Quality-of-Service (QoS) is satisfied and they can execute more computations to yield better results, if more system resources are available. These systems are typically considered under the Imprecise Computation (IC) model, achieving a better tradeoff between QoS and limited system resources. However, determining a QoS-aware mapping of these real-time IC-tasks onto the nodes of a CPS creates a set of interesting problems. In [18], we firstly propose a mathematical model to capture the dependency, energy and real-time constraints of IC-tasks, as well as the sensing, acting, and routing in the CPS. The problem is formulated as a Mixed-Integer Non-Linear Programming (MINLP) due to the complex nature of the problem. Secondly, to efficiently solve this problem, we provide a linearization method that results in a Mixed-Integer Linear Programming (MILP) formulation of our original problem. Finally, we decompose the transformed problem into a task allocation subproblem and a task adjustment subproblem, and, then, we find the optimal solution based on subproblem iteration. Through the simulations, we demonstrate the effectiveness of the proposed method. Last, but not least, wireless charging can provide dynamic power supply for CPS. Such systems are typically considered under the scenario of Wireless Rechargeable Sensor Networks (WRSNs). With the use of Mobile Chargers (MCs), the flexibility of WRSNs is further enhanced. However, the use of MCs poses several challenges during the system design. The coordination process has to simultaneously optimize the scheduling, the moving time and the charging time of multiple MCs, under limited system resources (e.g., time and energy). Efficient methods that jointly solve these challenges are generally lacking in the literature. In [17], we address the multiple MCs coordination problem under multiple system requirements. Firstly, we aim at minimizing the energy consumption of MCs, guaranteeing that every sensor will not run out of energy. We formulate the multiple MCs coordination problem as a mixed-integer linear programming and derive a set of desired network properties. Secondly, we propose a novel decomposition method to optimally solve the problem, as well as to reduce the computation time. Our approach divides the problem into a subproblem for the MC scheduling and a subproblem for the MC moving time and charging time, and solves them iteratively by utilizing the solution of one into the other. The convergence of the proposed method is analyzed theoretically. Simulation results demonstrate the effectiveness and scalability of the proposed method in terms of solution quality and computation time.

6.1.10. Fault-Tolerant Microarchitectures

Participants: Joseph Paturel, Angeliki Kritikakou, Olivier Sentieys.

As transistors scale down, processors are more vulnerable to radiation that can cause multiple transient faults in function units. Rather than excluding these units from execution, performance overhead of VLIW processors can be reduced when fault-free components of these affected units are still used. In [30], the function units are enhanced with coarse-grained fault detectors. A re-scheduling of the instructions is performed at run-time to use not only the healthy function units, but also the fault-free components of the faulty function units. The scheduling window of the proposed mechanism covers two instruction bundles, which makes it suitable to explore mitigation solutions in the current and in the next instruction execution. Experiments show that the proposed approach can mitigate a large number of faults with low performance and area overheads. In addition, technology scaling can cause transient faults with long duration. In this case, the affected function unit is usually considered as faulty and is not further used. To reduce this performance degradation, we proposed a hardware mechanism to (i) detect the faults that are still active during execution and (ii) re-schedule the instructions to use the fault-free components of the affected function units [31]. When the fault faints, the affected function unit components can be reused. The scheduling window of the proposed mechanism is two instruction bundles being able to exploit function units of both the current and the next instruction execution. The results show multiple long-duration fault mitigation can be achieved with low performance, area, and power overhead.

Simulation-based fault injection is commonly used to estimate system vulnerability. Existing approaches either partially model the studied system's fault masking capabilities, losing accuracy, or require prohibitive estimation times. Our work proposes a vulnerability analysis approach that combines gate-level fault injection with microarchitecture-level Cycle-Accurate and Bit-Accurate simulation, achieving low estimation time. Faults both in sequential and combinational logic are considered and fault masking is modeled at gate-level, microarchitecture-level and application-level, maintaining accuracy. Our case-study is a RISC-V processor. Obtained results show a more than 8% reduction in masked errors, increasing more than 55% system failures compared to standard fault injection approaches. This work is currently under review.

6.1.11. Fault-Tolerant Networks-on-Chip

Participants: Romain Mercier, Cédric Killian, Angeliki Kritikakou, Daniel Chillet.

Network-on-Chip has become the main interconnect in the multicore/manycore era since the beginning of this decade. However, these systems become more sensitive to faults due to transistor shrinking size. In parallel, approximate computing appears as a new computation model for applications since several years. The main characteristic of these applications is to support the approximation of data, both for computations and for communications. To exploit this specific application property, we develop a fault-tolerant NoC to reduce the impact of faults on the data communications. To address this problem, we consider multiple permanent faults on router which cannot be managed by Error-Correcting Codes (ECCs) and we propose a bit-shuffling method to reduce the impact of faults on Most Significant Bits (MSBs), hence permanent faults only impact Low Significant Bits (LSBs) instead of MSBs reducing the errors impact. We evaluated the proposed method for data mining benchmark and we show that our proposal can lead to 73.04% reduction on the clustering error rate and 84.64% reduction on the mean centroid Mean Square Error (MSE) for 3-bit permanent faults which affect MSBs on 32-bit words with a limited area cost. This work is currently under review for an international conference.

6.1.12. Improving the Reliability of Wireless Network-on-Chip (WiNoC)

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

Wireless Network-on-Chip (WiNoC) is one of the most promising solutions to overcome multi-hop latency and high power consumption of modern many/multi core System-on-Chip (SoC). However, standard WiNoC approaches are vulnerable to multi-path interference introduced by on-chip physical structures. To overcome such parasitic phenomenon, we first proposed a Time-Diversity Scheme (TDS) to enhance the reliability of on-chip wireless links using a realistic wireless channel model. We then proposed an adaptive digital transceiver, which enhances communication reliability under different wireless channel configurations in [39]. Based on the same realistic channel model, we investigated the impact of using some channel correction techniques. Experimental results show that our approach significantly improves Bit Error Rate (BER) under different wireless channel configurations. Moreover, our transceiver is designed to be adaptive, which allows for wireless communication links to be established in conditions where this would not be possible for standard transceiver architectures. The proposed architecture, designed using a 28-nm FDSOI technology, consumes only 3.27 mW for a data rate of 10 Gbit/s and has a very small area footprint. We also proposed a low-power, high-speed, multi-carrier reconfigurable transceiver based on Frequency Division Multiplexing (FDM) to ensure data transfer in future Wireless NoCs in [38]. The proposed transceiver supports a medium access control method to sustain unicast, broadcast and multicast communication patterns, providing dynamic data exchange among wireless nodes. Designed using a 28-nm FDSOI technology, the transceiver only consumes 2.37 mW and 4.82 mW in unicast/broadcast and multicast modes, respectively, with an area footprint of 0.0138 mm^2.

6.1.13. Error Mitigation in Nanophotonic Interconnect

Participants: Jaechul Lee, Cédric Killian, Daniel Chillet.

The energy consumption of manycore is dominated by data movements, which calls for energy-efficient and high-bandwidth interconnects. Integrated optics is promising technology to overcome the bandwidth limitations of electrical interconnects. However, it suffers from high power overhead related to low efficiency lasers, which calls for the use of approximate communications for error tolerant applications. In this context, in [26] we investigate the design of an Optical NoC supporting the transmission of approximate data. For this purpose, the least significant bits of floating point numbers are transmitted with low power optical signals. A transmission model allows estimating the laser power according to the targeted BER and a micro-architecture allows configuring, at run-time, the number of approximated bits and the laser output powers. Simulation results show that, compared to an interconnect involving only robust communications, approximations in the optical transmissions lead to a laser power reduction up to 42% for image processing application with a limited degradation at the application level.

6.2. Compilation and Synthesis for Reconfigurable Platform

6.2.1. Compile Time Simplification of Sparse Matrix Code Dependences

Participant: Tomofumi Yuki.

In [29], we developed a combined compile-time and runtime loop-carried dependence analysis of sparse matrix codes and evaluated its performance in the context of wavefront parallellism. Sparse computations incorporate indirect memory accesses such as x[col[j]] whose memory locations cannot be determined until runtime. The key contributions are two compile-time techniques for significantly reducing the overhead of runtime dependence testing: (1) identifying new equality constraints that result in more efficient runtime inspectors, and (2) identifying subset relations between dependence constraints such that one dependence test subsumes another one that is therefore eliminated. New equality constraints discovery is enabled by taking advantage of domain-specific knowledge about index arrays, such as col[j]. These simplifications lead to automatically-generated inspectors that make it practical to parallelize such computations. We analyze our simplification methods for a collection of seven sparse computations. The evaluation shows our methods reduce the complexity of the runtime inspectors significantly. Experimental results for a collection of five large matrices show parallel speedups ranging from 2x to more than 8x running on a 8-core CPU.

6.2.2. Study of Polynomial Scheduling

Participant: Tomofumi Yuki.

We have studied the Handelman's theorem used for polynomial scheduling, which resembles the Farkas' lemma for affine scheduling. Theorems from real algebraic geometry and polynomial optimization show that some polynomials have Handelman representations when they are non-negative on a domain, instead of strictly positive as stated in Handelman's theorem. The global minimizers of a polynomial must be at the boundaries of the domain to have such a representation with finite bounds on the degree of monomials. This creates discrepancies in terms of polynomials included in the exploration space with a fixed bound on the monomial degree. Our findings give an explanation to our failed attempt to apply polynomial scheduling to Index-Set Splitting: we were precisely trying to find polynomials with global minimizers at the interior of a domain.

6.2.3. Optimizing and Parallelizing compilers for Time-Critical Systems

Participant: Steven Derrien.

6.2.3.1. Contentions-Aware Task-Level Parallelization

Accurate WCET analysis for multicores is challenging due to concurrent accesses to shared resources, such as communication through bus or Network on Chip (NoC). Current WCET techniques either produce pessimistic WCET estimates or preclude conflicts by constraining the execution, at the price of a significant hardware under-utilization. Most existing techniques are also restricted to independent tasks, whereas real-time workloads will probably evolve toward parallel programs. The WCET behavior of such parallel programs is even more challenging to analyze because they consist of *dependent* tasks interacting through complex synchronization/communication mechanisms. In [36], we propose a scheduling technique that jointly selects Scratchpad Memory (SPM) contents off-line, in such a way that the cost of SPM loading/unloading is hidden. Communications are fragmented to augment hiding possibilities. Experimental results show the effectiveness of the proposed technique on streaming applications and synthetic task-graphs. The overlapping of communications with computations allows the length of generated schedules to be reduced by 4% on average on streaming applications, with a maximum of 16%, and by 8% on average for synthetic task graphs. We further show on a case study that generated schedules can be implemented with low overhead on a predictable multicore architecture (Kalray MPPA).

6.2.3.2. WCET-Aware Parallelization of Model-Based Applications for Multicores

Parallel architectures are nowadays increasingly used in embedded time-critical systems. The 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. The Argo toolchain operates on Scilab and XCoS inputs, and targets ScratchPad Memory (SPM)-based multicores. Data-layout and loop transformations play a key role in this flow as they improve SPM efficiency and reduce the number of accesses to shared main memory. In [20] we present the overall results of the project, a compiler tool-flow for automated parallelization of model-based real-time software, which addresses the shortcomings of multi-core architectures in real-time systems. The flow is demonstrated using a model-based Terrain Awareness and Warning Systems (TAWS) and an edge detection algorithm from the image-processing domain. Model-based applications are first transformed into real-time C code and from there into a well-predictable parallel C program. Tight bounds for the Worst-Case Execution Time (WCET) of the parallelized program can be determined using an integrated multicore WCET analysis. Thanks to the use of an architecture description language, the general approach is applicable to a wider range of target platforms. An experimental evaluation for a research architecture with network-on-chip (NoC) interconnect shows that the parallel WCET of the TAWS application can be improved by factor 1.77 using the presented compiler tools.

6.2.3.3. WCET oriented Iterative compilation

Static Worst-Case Execution Time (WCET) estimation techniques operate upon the binary code of a program in order to provide the necessary input for schedulability analysis techniques. Compilers used to generate this binary code include tens of optimizations, that can radically change the flow information of the program. Such information is hard to maintain across optimization passes and may render automatic extraction of important flow information, such as loop bounds, impossible. Thus, compiler optimizations, especially the sophisticated optimizations of mainstream compilers, are typically avoided. In this work, published in [23], we explore for the first time iterative-compilation techniques that reconcile compiler optimizations and static WCET estimation. We propose a novel learning technique that selects sequences of optimizations that minimize the WCET estimate of a given program. We experimentally evaluate the proposed technique using an industrial WCET estimation tool (AbsInt aiT) over a set of 46 benchmarks from four different benchmarks suites, including reference WCET benchmark applications, image processing kernels and telecommunication applications. Experimental results show that WCET estimates are reduced on average by 20.3% using the proposed technique,as compared to the best compiler optimization level applicable.

6.2.4. Towards Generic and Scalable Word-Length Optimization

Participants: Van-Phu Ha, Tomofumi Yuki, Olivier Sentieys.

Fixed-Point arithmetic is widely used for implementing Digital Signal Processing (DSP) systems on electronic devices. Since initial specifications are often written using floating-point arithmetic, conversion to fixed-point is a recurring step in hardware design. The primary objective of this conversion is to minimize the cost (energy and/or area) while maintaining an acceptable level of quality at the output. In Word-Length Optimization (WLO), each variable/operator may be assigned a different fixed-point encoding, which means that the design space grows exponentially as the number of variables increases. This is especially true when targeting hardware accelerators implemented in FPGA or ASIC. Thus, most approaches for WLO involve heuristic search algorithms. In [25] (a preliminary version also in [41]), we propose a method to improve the scalability of Word-Length Optimization (WLO) for large applications that use complex quality metrics such as Structural Similarity (SSIM). The input application is decomposed into smaller kernels to avoid uncontrolled explosion of the exploration time, which is known as noise budgeting. The main challenge addressed in this paper is how to allocate noise budgets to each kernel. This requires capturing the interactions across kernels. The main idea is to characterize the impact of approximating each kernel on accuracy/cost through simulation and regression. Our approach improves the scalability while finding better solutions for Image Signal Processor pipeline.

In [27], we propose an analytical approach to study the impact of floating-point (FlP) precision variation on the square root operation, in terms of computational accuracy and performance gain. We estimate the round-off error resulting from reduced precision. We also inspect the Newton Raphson algorithm used to approximate the square root in order to bound the error caused by algorithmic deviation. Consequently, the implementation of the square root can be optimized by fittingly adjusting its number of iterations with respect to any given FlP precision specification, without the need for long simulation times. We evaluate our error analysis of the square root operation as part of approximating a classic data clustering algorithm known as K-means, for the purpose of reducing its energy footprint. We compare the resulting inexact K-means to its exact counterpart, in the context of color quantization, in terms of energy gain and quality of the output. The experimental results show that energy savings could be achieved without penalizing the quality of the output (e.g., up to 41.87% of energy gain for an output quality, measured using structural similarity, within a range of [0.95,1]).

6.2.5. Optimized Implementations of Constant Multipliers for FPGAs

Participant: Silviu-Ioan Filip.

The multiplication by a constant is a frequently used arithmetic operation. To implement it on Field Programmable Gate Arrays(FPGAs), the state of the art offers two completely different methods: one relying on bit shifts and additions/subtractions, and another one using look-up tables and additions. So far, it was unclear which method performs best for a given constant and input/output data types. The main contribution of the work published in [40] is a thorough comparison of both methods in the main application contexts of constant multiplication: filters, signal-processing transforms, and elementary functions. Most of the previous state of the art addresses multiplication by an integer constant. This work shows that, in most of these application contexts, a formulation of the problem as the multiplication by a real constant allows for more efficient architectures. Another contribution is a novel extension of the shift-and-add method to real constants. For that, an integer linear programming (ILP) formulation is proposed, which truncates each component in the shift-and-add network to a minimum necessary word size that is aligned with the approximation error of the coefficient. All methods are implemented within the open-source FloPoCo framework.

6.2.6. Optimal Multiplierless FIR Filter Design

Participant: Silviu-Ioan Filip.

The hardware optimization of direct form finite impulse response (FIR) filters has been a topic of research for the better part of the last four decades and is still garnering significant research and industry interest. In [48], we present two novel optimization methods based on integer linear programming (ILP) that minimize the number of adders used to implement a direct/transposed FIR filter adhering to a given frequency specification. The proposed algorithms work by either fixing the number of adders used to implement the products (multiplier block adders) or by bounding the adder depth (AD) used for these products. The latter can be used to design filters with minimal AD for low power applications. In contrast to previous multiplierless FIR approaches, the methods introduced here ensure adder count optimality. To demonstrate their effectiveness, we perform several experiments using established design problems from the literature, showing superior results.

6.2.7. Application-specific arithmetic in high-level synthesis tools

Participant: Steven Derrien.

In [50], we have shown that the use of non-conventional implementation for floating-point arithmetic can bring significant benefits when used in the context of High-Level Synthesis. We are currently building on these preliminary results to show that it is possible to implement accelerators using exact floating-point arithmetic for similar performance/area cost than standard floating-point operators implementations. Our approach builds on Kulish's approach to implement floating-point adders, and targets dense Matrix Products kernels (GEM3 like) accelerators on FPGAs.

6.3. Applications

6.3.1. SmartSense

Participants: Nicolas Roux, Olivier Sentieys.

Developing smarter and greener buildings has been an expanding field of research over the last decades. One of the essential requirements for energy utilities is the knowledge of power consumption patterns at the single-appliance level. To estimate these patterns without using an individual power meter for each appliance, Non-Intrusive Load Monitoring (NILM) consists in disaggregating electrical loads by examining the appliance specific power consumption signature within the aggregated load single measurement. Therefore, the method is considered non-intrusive since the data are collected from a single electrical panel outside of the monitored building. Thus, NILM has been a very active field of research with renewed interest over the last years.

Therefore, knowing the plug-level power consumption of each appliance in a building can lead to drastic savings in energy consumption. In [35], we have addressed the issue of NILM inaccuracy in the context of industrial or commercial buildings, by combining data from a low-cost, general-purpose, wireless sensor network. We have proposed a novel approach based on a simplex solver to estimate the power load values of the steady states on sliding windows of data with varying size. We have shown the principle of the approach and demonstrated its interest, limited complexity, and ease of use.

7. Partnerships and Cooperations

7.1. Regional Initiatives

7.1.1. 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. In this project, CAIRN is addressing 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. For more details see https://bbc.cominlabs.u-bretagneloire.fr

7.2. National Initiatives

7.2.1. ANR AdequateDL

Participants: Olivier Sentieys, Silviu-Ioan Filip.

Program: ANR PRC

Project acronym: AdequateDL

Project title: Approximating Deep Learning Accelerators

Duration: Jan. 2019 - Dec. 2022

Coordinator: Cairn

Other partners: INL, CAIRN, LIRMM, CEA-LIST

The design and implementation of convolutional neural networks for deep learning is currently receiving a lot of attention from both industrials and academics. However, the computational workload involved with CNNs is often out of reach for low power embedded devices and is still very costly when run on datacenters. By relaxing the need for fully precise operations, approximate computing substantially improves performance and energy efficiency. Deep learning is very relevant in this context, since playing with the accuracy to reach adequate computations will significantly enhance performance, while keeping quality of results in a user-constrained range. AdequateDL will explore how approximations can improve performance and energy efficiency of hardware accelerators in deep-learning applications. Outcomes include a framework for accuracy exploration and the demonstration of order-of-magnitude gains in performance and energy efficiency of the proposed adequate accelerators with regards to conventional CPU/GPU computing platforms.

7.2.2. ANR RAKES

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

Program: ANR PRC

Project acronym: RAKES

Project title: Radio Killed an Electronic Star: speed-up parallel programming with broadcast communications based on hybrid wireless/wired network on chip

Duration: June 2019 - June 2023

Coordinator: TIMA

Other partners: TIMA, CAIRN, Lab-STICC

The efficient exploitation by software developers of multi/many-core architectures is tricky, especially when the specificities of the machine are visible to the application software. To limit the dependencies to the architecture, the generally accepted vision of the parallelism assumes a coherent shared memory and a few, either point to point or collective, synchronization primitives. However, because of the difference of speed between the processors and the main memory, fast and small dedicated hardware controlled memories containing copies of parts of the main memory (a.k.a caches) are used. Keeping these distributed copies up-to-date and synchronize the accesses to shared data, requires to distribute and share information between some may if not all the nodes. By nature, radio communications provide broadcast capabilities at negligible latency, they have thus the potential to disseminate information very quickly at the scale of a circuit and thus to be an opening for solving these issues. In the RAKES project, we intend to study how wireless communications can solve the scalability of the abovementioned problems, by using mixed wired/wireless Network on Chip. We plan to study several alternatives and to provide (a) a virtual platform for evaluation of the solutions and (b) an actual implementation of the solutions.

7.2.3. ANR Opticall^2

Participants: Olivier Sentieys, Cédric Killian, Daniel Chillet.

Program: ANR PRCE

Project acronym: Opticall^2

Project title: on-chip OPTIcal interconnect for ALL to ALL communications

Duration: Dec. 2018 - Nov. 2022

Coordinator: INL

Other partners: INL, CAIRN, C2N, CEA-LETI, Kalray

The aim of Opticall^2 is to design broadcast-enabled optical communication links in manycore architectures at wavelengths around 1.3μm. We aim to fabricate an optical broadcast link for which the optical power is equally shared by all the destinations using design techniques (different diode absorption lengths, trade-off depending on the current point in the circuit and the insertion losses). No optical switches will be used, which will allow the link latency to be minimized and will lead to deterministic communication times, which are both key features for efficient cache coherence protocols. The second main objective of Opticall^2 is to propose and design a new broadcast-aware cache coherence communication protocol allowing hundreds of computing clusters and memories to be interconnected, which is well adapted to the broadcast-enabled optical communication links. We expect better performance for the parallel execution of benchmark programs, and lower overall power consumption, specifically that due to invalidation or update messages.

7.2.4. ANR SHNOC

Participants: Cédric Killian, Daniel Chillet, Olivier Sentieys, Emmanuel Casseau.

Program: ANR JCJC (young researcher)

Project acronym: SHNOC

Project title: Scalable Hybrid Network-on-Chip

Duration: Feb. 2019 - Jan. 2022

P.I.: C. Killian, CAIRN

The goal of the SHNoC project is to tackle one of the manycore interconnect issues (scalability in terms of energy consumption and latency provided by the communication medium) by mixing emerging technologies. Technology evolution has allowed for the integration of silicon photonics and wireless on-chip communications, creating Optical and Wireless NoCs (ONoCs and WNoCs, respectively) paradigms. The recent publications highlight advantages and drawbacks for each technology: WNoCs are efficient for broadcast, ONoCs have low latency and high integrated density (throughput/cm^2) but are inefficient in multicast, while ENoCs are still the most efficient solution for small/average NoC size. The first contribution of this project is to study the compatibility of processes to associate the three aforementioned technologies and to define an hybrid topology of the interconnection architecture. This exploration will determine the number of antennas for the WNoC, the amount of embedded lasers sources for the ONoC and the routers architecture for the ENoC. The second main contribution is to provide quality of service of communication by determining, at run-time, the best path among the three NoCs with respect to a target, e.g. minimizing the latency or energy. We expect to demonstrate that the three technologies are more efficient when jointly used and combined, with respect to traffic characteristics between cores and quality of service targeted.

7.2.5. IPL ZEP

Participants: Davide Pala, Olivier Sentieys.

Program: Inria Project Lab

Project acronym: ZEP

Project title: Zero-Power Computing Systems

Duration: Oct. 2017 - Nov. 2020

Coordinator: Inria Socrate

Other partners: Pacap, Cairn, Corse, CEA-LETI

The ZEP project addresses the issue of designing tiny, batteryless, computing objects harvesting energy in the environment. The main application target is Internet of Things (IoT) where small communicating objects will be composed of this computing part associated to a low-power wake-up radio system. The energy level harvested being very low, very frequent energy shortages are expected, which makes the systems following the paradigm of Intermittently-Powered Systems. In order for the system to maintain a consistent state, it will be based on a new architecture embedding non-volatile memory (NVRAM). The major outcomes of the project will be a prototype harvesting board including NVRAM and the design of a new non-volatile processor (NVP) associated with its optimizing compiler and operating system. Cairn is focusing on the microarchitecture of the NVP and on new strategies for backup and restore data and processor state. The ZEP project gathers four Inria teams that have a scientific background in architecture, compilation, operating system and low power together with the CEA Grenoble. Another important goal of the project is to structure the research and innovation that should occur within Inria to prepare the important technological shift brought by NVRAM technologies.

7.2.6. DGA RAPID - FLODAM (2017–2021)

Participants: Joseph Paturel, Simon Rokicki, 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. For more details see https://flodam.fr

7.3. European Initiatives

7.3.1. H2020 ARGO

Participants: Steven Derrien, Angeliki Kritikakou, Olivier Sentieys.

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 (Germany), UR1/Inria/CAIRN, Recore Systems (Netherlands), TEI-WG (Greece), Scilab Ent. (France), Absint (Ger.), DLR (Ger.), Fraunhofer (Ger.)

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.

7.3.2. ANR International ARTEFaCT

Participants: Olivier Sentieys, Van-Phu Ha, 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, CAIRN, EPFL

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.

7.4. International Initiatives

7.4.1. Inria International Labs

EPFL-Inria

Associate Team involved in the International Lab:

7.4.1.1. IoTA

Title: Ultra-Low Power Computing Platform for IoT leveraging Controlled Approximation

International Partner (Institution - Laboratory - Researcher):

Ecole Polytechnique Fédérale de Lausanne (Switzerland) - Prof. 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 low-power 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.

7.4.2. Inria Associate Teams Not Involved in an Inria International Labs

7.4.2.1. IntelliVIS

Title: Design Automation for Intelligent Vision Hardware in Cyber Physical Systems

International Partner (Institution - Laboratory - Researcher):

IIT Goa (India) - Prof. Sharad Sinha

Start year: 2019

The proposed collaborative research work is focused on the design and development of artificial intelligence based embedded vision architectures for cyber physical systems (CPS). Embedded vision architectures for cyber physical systems (CPS), sometimes referred to as “Visual IoT”, are challenging to design because of primary constraints of compute resources, energy and power management. Embedded vision nodes in CPS, when designed with the application of Artificial Intelligence principles and algorithms, will turn into intelligent nodes (self-learning devices) capable of performing computation and inference at the node resulting in node-level cognition. This would allow only necessary and relevant post processed data to be sent to a human or a computer-based analyst for further processing and refinement in results. However, design and development of such nodes is non-trivial. Many existing computer vision algorithms, typically ported to embedded platforms, are compute and memory intensive thus limiting the operational time when ported to battery powered devices. In addition, transmission of captured visual data, with minimal processing at the node to extract actionable insights poses increased demands on computational, communication and energy requirements. Visual saliency i.e. extraction of key features or regions of interest in images or videos captured by an embedded vision node and related post processing for inference using AI techniques is an interesting and challenging research direction. The primary reason being that such an approach is expected to cover a wider range of application specific scenarios than statically determined approaches specific to each scenario involving remote off-loading of compute or scenario specific data on servers. Apart from a general approach to visual saliency in nodes using AI based methods (machine and deep learning methods), another principal goal of the proposed project is also to examine and propose methods that allow rapid deployment of AI techniques in these nodes. Many AI techniques are data driven and for a node to adapt from one environment or application specific scenario to another, rapid deployment of AI techniques over the air (OTA) would be an interesting and challenging research direction.

7.4.3. Inria International Partners

7.4.3.1. DARE

Title: Design space exploration Approaches for Reliable Embedded systems

International Partner (Institution - Laboratory - Researcher):

IMEC (Belgium) - Francky Catthoor, IMEC fellow

Duration: 2017 - 2021

Start year: 2017

This collaborative research focuses on methodologies to design low cost and efficient techniques for safety-critical embedded systems, which require high performance and safety implying both fault tolerance and hard real-time constraints. More precisely, the objective is to develop Design Space Exploration (DSE) methodology applicable to any platform domain to drive the design of adaptive predictable low cost and efficient error detection techniques. Run-time dynamic control mechanisms are proposed to actively optimize system fault tolerance by exploring the trade-offs between predictability, reliability, performance and energy consumption using the information received from the environment and the platform during execution. In contrast to design-time static approaches the dynamism can then be exploited to improve energy consumption and performance.

7.4.3.2. 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

7.4.3.3. HARAMCOP

Title: Hardware accelerators modeling using constraint-based programming

International Partner (Institution - Laboratory - Researcher):

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

7.4.3.4. DeLeES

Title: Energy-efficient Deep Learning Systems for Low-cost Embedded Systems

International Partner (Institution - Laboratory - Researcher):

University of British Columbia (Vancouver, Canada) - Electrical and Computer Engineering - Prof. Guy Lemieux

Start year: 2018

This collaboration is centered around creation of deep-learning inference systems which are energy efficient and low cost. There are two design approaches: (i) an all-digital low-precision system, and (ii) mixed analog/digital low-precision system.

7.4.3.5. Informal International Partners

Dept. of Electrical and Computer Engineering, Concordia University (Canada), Optical network-on-chip, manycore architectures.

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.

PARC Lab., the University of Auckland (New-Zealand), Fault-tolerant task scheduling onto multi-core.

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.

7.5. International Research Visitors

7.5.1. Visits of International Scientists

- Bernard Goossens, Univ. Perpignan, July 2019.
- Sharad Sinha, IIT Goa, India, July 2019.

7.5.2. Visits to International Teams

7.5.2.1. Sabbatical programme

Steven Derrien visited Colorado State University for a 6 month sabbatical from January to July 2019, where he collaborated with Sanjay Rajopadhye. This collaboration has led to two joint PhD between Université de Rennes 1 and Colorado State University which both started in late 2019.

7.5.2.2. Research Stays Abroad

- Olivier Sentieys visited Colorado State University, Computer Science Department and gave a seminar on Approximate Computing in November 2019.
- P. Dobias (PhD student) spent 5 months in the Parallel and Reconfigurable Lab. of the Electrical and Computer Engineering department, the University of Auckland, New Zealand, from November 2018 until March 2019.

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Scientific Events: Organisation

8.1.1.1. General Chair, Scientific Chair

- D. Chillet was the General Co-Chair of HiPEAC RAPIDO'19 Workshop.

8.1.1.2. Member of the Organizing Committees

- E. Casseau is a member of DASIP Steering Committee, Conference on Design and Architectures for Signal and Image Processing.

8.1.2. Scientific Events: Selection

8.1.2.1. Chair of Conference Program Committees

- O. Sentieys is Co-Chair of the D8 Track on Architectural and Microarchitectural Design at IEEE/ACM DATE since 2018.
- O. Sentieys is a member of the committee for delivering the Best Paper Award at IEEE/ACM DATE 2020.
- O. Sentieys served as a committee member in the IEEE EDAA Outstanding Dissertations Award (ODA).

8.1.2.2. 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'19, IEEE FPT'19, IEEE ASAP'19 and ARC'19.
- A. Kritikakou was a member of technical program committee of IEEE RTAS'20, ECRTS'19, SAMOS'19, DATE'20.
- O. Sentieys was a member of technical program committee of IEEE/ACM DATE, IEEE FPL, ACM ENSSys, ACM SBCCI, IEEE ReConFig.
- T. Yuki was a member of technical program committee of CGO '19, SC '19, TAPAS '19, CC '20, IMPACT '20, and was a member of external review committee of PACT '19.

8.1.3. Journal

8.1.3.1. Member of the Editorial Boards

- 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.

8.1.4. Invited Talks

- O. Sentieys gave an invited talk at FETCH (École d'hiver Francophone sur les Technologies de Conception des Systèmes embarqués Hétérogènes), Louvain-la-Neuve, Belgique, January 2019 on "Approximating Deep Learning Accelerators".

- O. Sentieys gave an invited talk at ARCHI Spring School, Lorient, France, May 2019 on “Design of VLSI Integrated Circuits - A (very) deep dive into computing chips”.
- O. Sentieys gave a Keynote at the IEEE International Nanodevices and Computing (INC) - IEEE International Conference on Rebooting Computing (ICRC), Grenoble, France, April 2019 on “Playing with numbers for Energy Efficiency: Introduction to Approximate Computing” [21].
- O. Sentieys gave a tutorial at the 22nd IEEE/ACM Design, Automation and Test in Europe (DATE), March 2019 on “A Comprehensive Analysis of Approximate Computing Techniques: From Component- to Application-Level” [47].
- D.Chillet gave a talk during GDR SoC2 and RO topic day, in november 2019 on “Mathematics model for wavelength allocation in the context of Optical NoC”.
- D. Chillet gave an invited talk at FETCH (École d’hiver Francophone sur les Technologies de Conception des Systèmes embarqués Hétérogènes), Louvain-la-Neuve, Belgique, January 2019 on ”Power management for communications on Optical NoC”.

8.1.5. Leadership within the Scientific Community

- E. Casseau is a member of the French National University Council in Signal Processing and Electronics (CNU - Conseil National des Universites, 61ème section) since 2018.
- D.Chillet is a member of the French National University Council in Signal Processing and Electronics (CNU - Conseil National des Universites, 61ème section) since 2019.
- D. Chillet is member of the Board of Directors of Gretsi Association.
- D. Chillet is co-animator of the "Connected Objects" topic of GDR SoC^2.
- 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).
- O. Sentieys is a member of the steering committee of GDR SoC^2.
- O. Sentieys is an elected member of the Evaluation Committee (CE) of Inria.

8.1.6. Scientific Expertise

- O. Sentieys served as an expert for Fund for Scientific Research – FNRS, Belgium.
- D.Chillet served as an expert for the call CAPES-COFECUB 2020.
- D.Chillet served as an expert for German Research Foundation.

8.2. Teaching - Supervision - Juries

8.2.1. Teaching Responsibilities

- E. Casseau is in charge of the Department of “Digital Systems” at ENSSAT Engineering Graduate School.
- D. Chillet was the responsible of the ICT Master of University of Science and Technology of Hanoi.
- C. Killian was the responsible of the second year of the ”Instrumentation” DUT at IUT, Lannion until July 2019.
- O. Sentieys is responsible of the ”Embedded Systems” major of the SISEA Master by Research.
- C. Wolinski was the Director of ESIR until May 2019.

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.

8.2.2. Teaching

E. Casseau: signal processing, 21h, ENSSAT (L3)

E. Casseau: low power design, 6h, ENSSAT (M1)

E. Casseau: real time design methodology, 57h, ENSSAT (M1)

E. Casseau: computer architecture, 24h, ENSSAT (M1)

E. Casseau: VHDL design, 42h, ENSSAT (M1)

E. Casseau: SoC and high-level synthesis, 33h, Master by Research (SISEA) and ENSSAT (M2)

S. Derrien, optimizing and parallelising compilers, 14h, Master of Computer Science, ISTIC(M2)

S. Derrien, advanced processor architectures, 8h, Master of Computer Science, ISTIC(M2)

S. Derrien, high level synthesis, 20h, Master of Computer Science, ISTIC(M2)

S. Derrien, computer science research projects, 10h, Master of Computer Science, ISTIC(M1)

S. Derrien: introduction to operating systems, 8h, ISTIC (M1)

S. Derrien, principles of digital design, 20h, Bachelor of EE/CS, ISTIC(L2)

S. Derrien, computer architecture, 48h, Bachelor of Computer Science, ISTIC(L3)

S.I. Filip, Operating Systems, 24h, Master of Mechatronics, ENS RENNES (M2)

F. Charot: computer architecture, 16h, ESIR (L3)

F. Charot: Computer architecture, 58h, ISTIC (L3)

D. Chillet: embedded processor architecture, 20h, ENSSAT (M1)

D. Chillet: multimedia processor architectures, 24h, ENSSAT (M2)

D. Chillet: advanced processor architectures, 20h, ENSSAT (M2)

D. Chillet: micro-controller, 64h, ENSSAT (L3)

D. Chillet: low-power digital CMOS circuits, 6h, Telecom Bretagne (M2)

D. Chillet: low-power digital CMOS circuits, 4h, UBO (M2)

C. Killian: digital electronics, 52h, IUT Lannion (L1)

C. Killian: automated measurements, 44h, 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, 60h, ISTIC (L3)

O. Sentieys: VLSI integrated circuit design, 24h, ENSSAT (M1)

O. Sentieys: VHDL and logic synthesis, 18h, 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)

8.2.3. Supervision

PhD: Audrey Lucas, Software support resistant to passive and active attacks for asymmetric cryptography on (very) small computation cores, Dec. 2019, A. Tisserand.

PhD: Genevieve Ndour, Approximate computing for high energy-efficiency in internet-of-things applications, Jul. 2019, A. Tisserand, A. Molnos (CEA LETI).

PhD in progress: Thibault Allenet, Low-Cost Neural Network Algorithms and Implementations for Temporal Sequence Processing, March 2019, O. Sentieys, O. Bichler (CEA LIST).

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

PhD in progress: Minyu Cui, Energy-Quality-Time Fault Tolerant Task Mapping on Multicore Architectures, Oct. 2018, E. Casseau, A. Kritikakou.

PhD in progress: Petr Dobias, Energy-Quality-Time Fault Tolerant Task Mapping on Multicore Architectures, Oct. 2017, E. Casseau.

PhD in progress: Corentin Ferry, Compiler support for Runtime data compression for FPGA accelerators, Sep. 2019, S. Derrien, T. Yuki and S. Rajopadhye (co-tutelle between Université de Rennes 1 and Colorado State University).

PhD in progress: Adrien Gaonac'h, Test de robustesse des systèmes embarqués par perturbation contrôlée en simulation à partir de plateformes virtuelles, Oct. 2019, D. Chillet, Yves Lhuillier (CEA LIST), Youri Helen (DGA).

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: Jaechul Lee, Energy-Performance Trade-Off in Optical Network-on-Chip, Dec. 2018, D. Chillet, C. Killian.

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

PhD in progress: Romain Mercier, Fault Tolerant Network on Chip for Deep Learning Algorithms, Oct. 2018, D. Chillet, C. Killian, A. Kritikakou.

PhD in progress: Louis Narmour, Revisiting memory allocation in tyeh polyhedral model, Sep. 2019, S. Derrien, T. Yuki and S. Rajopadhye (co-tutelle between Université de Rennes 1 and Colorado State University).

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: Davide Pala, Non-Volatile Processors for Intermittently-Powered Computing Systems, Jan. 2018, O. Sentieys, I. Miro-Panades (CEA LETI).

PhD in progress: Joseph Paturel, Design-space exploration of fault-tolerant multicores, Sep. 2018, O. Sentieys, A. Kritikakou.

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 (IRISA/GRANIT).

8.3. Popularization

8.3.1. Articles and contents

O. Sentieys contributed to the Inria white book on scientific reviews and challenges in cybersecurity https://files.inria.fr/dircom/extranet/LB_cybersecurity_WEB.pdf.

8.3.2. Interventions

Members of the team participated in the national science festival *(Fête de la Science)* in Lannion, October, with demonstrations on wireless sensor networks, body sensor network, cryptology and digital circuit design.

9. Bibliography

Major publications by the team in recent years

[1] R. DAVID, S. PILLEMENT, O. SENTIEYS.*Energy-Efficient Reconfigurable Processsors*, 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] C. HURIAUX, A. COURTAY, O. SENTIEYS.*Design Flow and Run-Time Management for Compressed FPGA Configurations*, in "IEEE/ACM Design, Automation and Test in Europe (DATE)", March 2015, https://hal.inria.fr/hal-01089319

[4] 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^{o} 4, p. 581-597 [*DOI :* 10.1007/s10270-012-0266-8], https://hal.inria.fr/hal-00717219

[5] 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^{o} 11, p. 1454-1464

[6] 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^{o} 2, p. 1-38, http://doi.acm.org/10.1145/2209285.2209289

[7] 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

[8] D. MENARD, O. SENTIEYS.*Automatic Evaluation of the Accuracy of Fixed-point Algorithms*, in "IEEE/ACM Design, Automation and Test in Europe (DATE-02)", Paris, March 2002

[9] 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

[10] 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^{o} 10, p. 2326 - 2339 [*DOI :* 10.1109/TCSI.2012.2188938], http://hal.inria.fr/hal-00741741

[11] C. WOLINSKI, M. GOKHALE, K. MCCABE.*A polymorphous computing fabric*, in "IEEE Micro", 2002, vol. 22, n^{o} 5, p. 56–68

[12] 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^{o} 1, p. 1–36, http://doi.acm.org/10.1145/1640457.1640458

Publications of the year

Doctoral Dissertations and Habilitation Theses

[13] A. LUCAS.*Robust software support against passive and active attacks for the arithmetic of asymmetric cryptography on (very) small computing cores*, Université Rennes 1, December 2019, https://hal.archives-ouvertes.fr/tel-02425706

[14] G. NDOUR.*Approximate computing for high energy-efficiency in internet-of-things applications*, Université Rennes 1, July 2019, https://hal.archives-ouvertes.fr/tel-02292988

Articles in International Peer-Reviewed Journal

[15] S.-I. FILIP, A. JAVEED, L. N. TREFETHEN.*Smooth random functions, random ODEs, and Gaussian processes*, in "SIAM Review", February 2019, vol. 61, n^o^ 1, p. 185-205 [*DOI :* 10.1137/17M1161853], https://hal.inria.fr/hal-01944992

[16] G. GALLIN, A. TISSERAND.*Generation of Finely-Pipelined GF(P) Multipliers for Flexible Curve based Cryptography on FPGAs*, in "IEEE Transactions on Computers", November 2019, vol. 68, n^o^ 11, p. 1612-1622 [*DOI :* 10.1109/TC.2019.2920352], https://hal.archives-ouvertes.fr/hal-02141260

[17] L. MO, A. KRITIKAKOU, S. HE.*Energy-Aware Multiple Mobile Chargers Coordination for Wireless Rechargeable Sensor Networks*, in "IEEE internet of things journal", May 2019, p. 1-13 [*DOI :* 10.1109/JIOT.2019.2918837], https://hal.inria.fr/hal-02140283

[18] L. MO, A. KRITIKAKOU.*Mapping Imprecise Computation Tasks on Cyber-Physical Systems*, in "Peer-to-Peer Networking and Applications", 2019, p. 1726–1740 [*DOI :* 10.1007/s12083-019-00749-9], https://hal.archives-ouvertes.fr/hal-02397099

[19] L. MO, P. YOU, X. CAO, Y.-Q. SONG, A. KRITIKAKOU.*Event-Driven Joint Mobile Actuators Scheduling and Control in Cyber-Physical Systems*, in "IEEE Transactions on Industrial Informatics", March 2019, p. 1-13 [*DOI :* 10.1109/TII.2019.2906061], https://hal.inria.fr/hal-02080647

[20] S. REDER, F. KEMPF, H. BUCHER, J. BECKER, P. ALEFRAGIS, N. S. VOROS, S. SKALISTIS, S. DERRIEN, I. PUAUT, O. OEY, T. STRIPF, C. FERDINAND, C. DAVID, P. ULBIG, D. MUELLER, U. DURAK.*Worst-Case Execution-Time-Aware Parallelization of Model-Based Avionics Applications*, in "Journal of Aerospace Information Systems", November 2019, vol. 16, n^o^ 11, p. 521-533 [*DOI :* 10.2514/1.I010749], https://hal.archives-ouvertes.fr/hal-02383381

Invited Conferences

[21] O. SENTIEYS.*Playing with number for Energy Efficiency, Introduction to Approximate Computing*, in "INC 2019 - IEEE International Nanodevices and Computing", Grenoble, France, IEEE, April 2019, https://hal.inria.fr/hal-02183527

International Conferences with Proceedings

[22] T. CONG, F. CHAROT.*Designing Application-Specific Heterogeneous Architectures from Performance Models*, in "MCSoC 2019 - IEEE 13th International Symposium on Embedded Multicore/Many-core Systems-on-Chip", Singapore, Singapore, October 2019, p. 1-8, https://hal.inria.fr/hal-02289868

[23] M. DARDAILLON, S. SKALISTIS, I. PUAUT, S. DERRIEN.*Reconciling Compiler Optimizations and WCET Estimation Using Iterative Compilation*, in "RTSS 2019 - 40th IEEE Real-Time Systems Symposium", Hong Kong, China, IEEE, December 2019, p. 1-13, https://hal.archives-ouvertes.fr/hal-02286164

[24] M. GUEGUEN, O. SENTIEYS, A. TERMIER.*Accelerating Itemset Sampling using Satisfiability Constraints on FPGA*, in "DATE 2019 - 22nd IEEE/ACM Design, Automation and Test in Europe", Florence, Italy, IEEE, March 2019, p. 1046-1051 [*DOI :* 10.23919/DATE.2019.8714932], https://hal.inria.fr/hal-01941862

[25] V.-P. HA, T. YUKI, O. SENTIEYS.*Towards Generic and Scalable Word-Length Optimization*, in "IEEE/ACM Design Automation and Test in Europe (DATE)", Grenoble, France, March 2020, https://hal.inria.fr/hal-02387232

[26] J. LEE, C. KILLIAN, S. LE BEUX, D. CHILLET.*Approximate nanophotonic interconnects*, in "NOCS 2019 - 13th IEEE/ACM International Symposium on Networks-on-Chip", New York, United States, ACM, October 2019, p. 1-7 [*DOI :* 10.1145/3313231.3352365], https://hal.archives-ouvertes.fr/hal-02341667

[27] O. MATOUSSI, Y. DURAND, O. SENTIEYS, A. MOLNOS.*Error Analysis of the Square Root Operation for the Purpose of Precision Tuning: a Case Study on K-means*, in "ASAP 2019 - 30th IEEE International Conference on Application-specific Systems, Architectures and Processors", New York, United States, IEEE, July 2019, p. 1-8, https://hal.inria.fr/hal-02183945

[28] L. MO, A. KRITIKAKOU, O. SENTIEYS.*Approximation-aware Task Deployment on Asymmetric Multicore Processors*, in "DATE 2019 - 22nd IEEE/ACM Design, Automation and Test in Europe", Florence, Italy, IEEE, March 2019, p. 1513-1518 [*DOI :* 10.23919/DATE.2019.8715077], https://hal.inria.fr/hal-01940358

[29] M. S. MOHAMMADI, M. M. STROUT, T. YUKI, K. CHESHMI, E. DAVIS, M. HALL, M. M. DEHNAVI, P. NANDY, C. OLSCHANOWSKY, A. VENKAT.*Sparse computation data dependence simplification for efficient compiler-generated inspectors*, in "PLDI 2019 - 40th ACM SIGPLAN Conference on Programming Language Design and Implementation", Phoenix, United States, ACM Press, November 2019, p. 594-609 [*DOI :* 10.1145/3314221.3314646], https://hal.inria.fr/hal-02396761

[30] R. PSIAKIS, A. KRITIKAKOU, O. SENTIEYS, E. CASSEAU.*Run-time Coarse-Grained Hardware Mitigation for Multiple Faults on VLIW Processors*, in "DASIP 2019 - Conference on Design and Architectures for Signal and Image Processing", Montréal, Canada, October 2019, p. 1-6, https://hal.inria.fr/hal-02344282

[31] R. PSIAKIS, A. KRITIKAKOU, O. SENTIEYS.*Fine-Grained Hardware Mitigation for Multiple Long-Duration Transients on VLIW Function Units*, in "DATE 2019 - 22nd IEEE/ACM Design, Automation and Test in Europe", Florence, Italy, IEEE, March 2019, p. 976-979 [*DOI :* 10.23919/DATE.2019.8714899], https://hal.inria.fr/hal-01941860

[32] S. ROKICKI, D. PALA, J. PATUREL, O. SENTIEYS.*What You Simulate Is What You Synthesize: Designing a Processor Core from C++ Specifications*, in "ICCAD 2019 - 38th IEEE/ACM International Conference on Computer-Aided Design", Westminster, CO, United States, IEEE, November 2019, p. 1-8, https://hal.archives-ouvertes.fr/hal-02303453

[33] S. ROKICKI, E. ROHOU, S. DERRIEN.*Aggressive Memory Speculation in HW/SW Co-Designed Machines*, in "DATE 2019 - 22nd IEEE/ACM Design, Automation and Test in Europe", Florence, Italy, IEEE, March 2019, p. 332-335 [*DOI :* 10.23919/DATE.2019.8715010], https://hal.archives-ouvertes.fr/hal-01941876

[34] S. ROKICKI.*GhostBusters: Mitigating Spectre Attacks on a DBT-Based Processor*, in "DATE 2020 - 23rd IEEE/ACM Design, Automation and Test in Europe", Grenoble, France, March 2020, https://hal.archives-ouvertes.fr/hal-02396631

[35] N. ROUX, B. VRIGNEAU, O. SENTIEYS.*Improving NILM by Combining Sensor Data and Linear Programming*, in "SAS 2019 - IEEE Sensors Applications Symposium", Sophia Antipolis, France, IEEE, March 2019, p. 1-6 [*DOI :* 10.1109/SAS.2019.8706021], https://hal.inria.fr/hal-02394920

[36] B. ROUXEL, S. SKALISTIS, S. DERRIEN, I. PUAUT.*Hiding Communication Delays in Contention-Free Execution for SPM-Based Multi-Core Architectures*, in "ECRTS 2019 - 31st Euromicro Conference on Real-Time Systems", Stuttgart, Germany, July 2019, p. 1-24 [*DOI :* 10.4230/LIPICS.ECRTS.2019.25], https://hal.archives-ouvertes.fr/hal-02190271

[37] S. SKALISTIS, A. KRITIKAKOU.*Timely Fine-grained Interference-sensitive Run-time Adaptation of Time-triggered Schedules*, in "RTSS 2019 - 40th IEEE Real-Time Systems Symposium", Hong Kong, China, IEEE, December 2019, p. 1-13, https://hal.archives-ouvertes.fr/hal-02316392

[38] J. O. SOSA, O. SENTIEYS, C. ROLAND, C. KILLIAN.*Multi-Carrier Spread-Spectrum Transceiver for WiNoC*, in "NOCS 2019 - 13th IEEE/ACM International Symposium on Networks-on-Chip", New York, United States, ACM, October 2019, p. 1-2 [*DOI :* 10.1145/3313231.3352373], https://hal.inria.fr/hal-02394890

[39] J. O. SOSA, O. SENTIEYS, C. ROLAND.*Adaptive Transceiver for Wireless NoC to Enhance Multicast/Unicast Communication Scenarios*, in "ISVLSI 2019 - IEEE Computer Society Annual Symposium on VLSI", Miami, United States, IEEE, July 2019, p. 1-6 [*DOI :* 10.1109/ISVLSI.2019.00111], https://hal.inria.fr/hal-02394902

[40] F. DE DINECHIN, S.-I. FILIP, L. FORGET, M. KUMM.*Table-Based versus Shift-And-Add constant multipliers for FPGAs*, in "ARITH 2019 - 26th IEEE Symposium on Computer Arithmetic", Kyoto, Japan, IEEE, June 2019, p. 1-8, https://hal.inria.fr/hal-02147078

Conferences without Proceedings

[41] V.-P. HA, T. YUKI, O. SENTIEYS.*Noise Budgeting in Multiple-Kernel Word-Length Optimization*, in "AxC 2019 - 4th Workshop on Approximate Computing", Florence, Italy, March 2019, p. 1-3, https://hal.inria.fr/hal-02183936

[42] S. ROKICKI, D. PALA, J. PATUREL, O. SENTIEYS.*What You Simulate Is What You Synthesize: Design of a RISC-V Core from C++ Specifications*, in "RISC-V Workshop 2019", Zurich, Switzerland, June 2019, p. 1-2, https://hal.inria.fr/hal-02394911

[43] B. ROUX, M. GAUTIER, O. SENTIEYS.*Exploration architecturale d'accélérateur pour des architectures multi-coeurs hétérogènes*, in "27ème colloque du Groupement de Recherche en Traitement du Signal et des Images", Lille, France, August 2019, https://hal.archives-ouvertes.fr/hal-02406976

[44] T. YUKI.*The Limit of Polynomials: Implications of Handelman's Theorem for Exploring Schedules*, in "IMPACT 2019 - 9th International Workshop on Polyhedral Compilation Techniques", Valencia, Spain, January 2019, p. 1-8, https://hal.inria.fr/hal-02397043

Scientific Books (or Scientific Book chapters)

[45] D. MENARD, G. CAFFARENA, J. A. LOPEZ, D. NOVO, O. SENTIEYS.*Fixed-point refinement of digital signal processing systems*, in "Digitally Enhanced Mixed Signal Systems", The Institution of Engineering and Technology, May 2019, n^o Chapter 1, p. 1-37 [*DOI :* 10.1049/PBCS040E_CH], https://hal.inria.fr/hal-01941898

[46] D. MÉNARD, G. CAFFARENA, J. A. LOPEZ, D. NOVO, O. SENTIEYS.*Analysis of Finite Word-Length Effects in Fixed-Point Systems*, in "Handbook of Signal Processing Systems", S. S. BHATTACHARYYA (editor), 2019, p. 1063-1101 [*DOI :* 10.1007/978-3-319-91734-4_29], https://hal.inria.fr/hal-01941888

Other Publications

[47] A. BOSIO, D. MENARD, O. SENTIEYS.*A Comprehensive Analysis of Approximate Computing Techniques: From Component- to Application-Level*, March 2019, p. 1-5, DATE 2019 - 22nd IEEE/ACM Design, Automation and Test in Europe, https://hal.inria.fr/hal-01941757

[48] M. KUMM, A. VOLKOVA, S.-I. FILIP.*Design of Optimal Multiplierless FIR Filters*, December 2019, https://arxiv.org/abs/1912.04210 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02392522

[49] S. ROKICKI, E. ROHOU, S. DERRIEN.*Hybrid-DBT: Hardware Accelerated Dynamic Binary Translation*, June 2019, 1, RISC-V 2019 - Workshop Zurich, Poster, https://hal.archives-ouvertes.fr/hal-02155019

[50] Y. UGUEN, F. DE DINECHIN, V. LEZAUD, S. DERRIEN.*Application-specific arithmetic in high-level synthesis tools*, December 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02423363

References in notes

[51] S. HAUCK, A. DEHON (editors). *Reconfigurable Computing: The Theory and Practice of FPGA-Based Computation*, Morgan Kaufmann, 2008

[52] 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^o 2, p. 167–184

[53] 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

[54] C. BOBDA.*Introduction to Reconfigurable Comp.: Architectures Algorithms and Applications*, Springer, 2007

[55] S. BORKAR, A. A. CHIEN.*The Future of Microprocessors*, in "Commun. ACM", May 2011, vol. 54, n^o 5, p. 67–77, http://doi.acm.org/10.1145/1941487.1941507

[56] 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

[57] K. COMPTON, S. HAUCK.*Reconfigurable computing: a survey of systems and software*, in "ACM Comput. Surv.", 2002, vol. 34, n^o 2, p. 171–210, http://doi.acm.org/10.1145/508352.508353

[58] 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

[59] 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^{o} 10, p. 1432- 1442

[60] 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

[61] R. H. DENNARD, F. H. GAENSSLEN, V. L. RIDEOUT, E. BASSOUS, A. R. LEBLANC.*Design of ion-implanted MOSFET's with very small physical dimensions*, in "IEEE Journal of Solid-State Circuits", 1974, vol. 9, n^{o} 5, p. 256–268

[62] 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

[63] 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

[64] 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^{o} 1, p. 26-33, http://doi.ieeecomputersociety.org/10.1109/MDT.2003.1173050

[65] 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

[66] 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

[67] 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

[68] 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

[69] A. PUTNAM, A. CAULFIELD, E. CHUNG, D. CHIOU, K. CONSTANTINIDES, J. DEMME, H. ESMAEILZADEH, 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

[70] G. THEODORIDIS, D. SOUDRIS, S. VASSILIADIS.*2*, in "A survey of coarse-grain reconfigurable architectures and CAD tools", Springer Verlag, 2007

[71] 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^{o} 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

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
A2.1.3. - Object-oriented programming
A2.1.4. - Functional programming
A2.1.12. - Dynamic languages
A2.2. - Compilation
A2.2.1. - Static analysis
A2.2.2. - Memory models
A2.2.3. - Memory management
A2.2.5. - Run-time systems
A2.2.9. - Security by compilation
A2.4. - Formal method for 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
A7.2.2. - Automated Theorem Proving
A7.2.3. - Interactive Theorem Proving

Other Research Topics and Application Domains:

B6.1. - Software industry
B6.1.1. - Software engineering
B6.4. - Internet of things
B6.6. - Embedded systems
B9.10. - Privacy

1. Team, Visitors, External Collaborators

Research Scientists

Thomas Jensen [Team leader, Inria, Senior Researcher, HDR]
Frédéric Besson [Inria, Researcher]
Alan Schmitt [Inria, Senior Researcher, HDR]
Benoît Montagu [Inria, Starting Research position, from Nov 2019]

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]

David Pichardie [Ecole normale supérieure de Rennes, Professor, HDR]

External Collaborator

Mickael Delahaye [DGA, from Nov 2019]

Technical Staff

Nicolas Barré [Ecole normale supérieure de Rennes, Engineer]
Benoît Montagu [Inria, Engineer, from May 2019 until Oct 2019]
Samuel Risbourg [Inria, Engineer]

PhD Students

Guillaume Ambal [Univ de Rennes I, PhD Student, from Sep 2019]
Aurele Barriere [Ecole Normale Supérieure Rennes, PhD Student, from Sep 2019]
Alexandre Dang [Inria, PhD Student, until Sep 2019]
Samy Daoud [Orange Labs, PhD Student, granted by CIFRE]
Timothée Haudebourg [Univ de Rennes I, PhD Student]
Rémi Hutin [Ecole normale supérieure de Rennes, PhD Student]
Aurèle Barrière [Ecole normale supérieure de Rennes, PhD Student]
Solène Mirliaz [Ecole normale supérieure de Rennes, PhD Student]
Adam Khayam [Inria, PhD Student, from Jul 2019]
Julien Lepiller [Inria, PhD Student]
Louis Noizet [Univ de Rennes 1, PhD Student, from Oct 2019]

Post-Doctoral Fellows

Stefania Dumbrava [Ecole normale supérieure de Rennes, Post-Doctoral Fellow, until Aug 2019]
Jean Christophe Lechenet [Ecole normale supérieure de Rennes, Post-Doctoral Fellow]
Thomas Rubiano [Univ de Rennes I, Post-Doctoral Fellow]

Visiting Scientist

Nathanael Courant [Ecole Normale Supérieure Paris, from Mar 2019 until Aug 2019]

2. Overall Objectives

2.1. Project overview

The overall goal of the CELTIQUE project is to improve the security and reliability of software with semantics-based 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, 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. CompcertSSA

KEYWORDS: Optimizing compiler - Formal methods - Proof assistant - SSA

FUNCTIONAL DESCRIPTION: CompcertSSA is built on top of the Compcert verified C compiler, by adding a middle-end based on the SSA form (Static Single Assignment) : conversion to SSA, SSA-based optimizations, and destruction of SSA.

- Participants: Sandrine Blazy, Delphine Demange, Yon Fernandez de Retana, David Pichardie and Leo Stefanesco
- Contact: Delphine Demange
- Publications: Mechanizing conventional SSA for a verified destruction with coalescing - Validating Dominator Trees for a Fast, Verified Dominance Test - A Formally Verified SSA-based Middle-end : Static Single Assignment meets CompCert - Formal Verification of an SSA-based Middle-end for CompCert - Verifying Fast and Sparse SSA-based Optimizations in Coq.
- URL: http://compcertssa.gforge.inria.fr/

3.2. 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

3.3. Javalib

KEYWORDS: Library - Java - Ocaml

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: David Pichardie
- URL: http://sawja.inria.fr/

3.4. 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/

3.5. 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: David Pichardie
- URL: http://sawja.inria.fr/

3.6. Timbuk

KEYWORDS: Automated deduction - Ocaml - Program verification - Tree Automata - Term Rewriting Systems

FUNCTIONAL DESCRIPTION: Timbuk is a tool designed to compute or over-approximate sets of terms reachable by a given term rewriting system. The libray also provides an OCaml toplevel with all usual functions on Bottom-up Nondeterministic Tree Automata.

- Participant: Thomas Genet
- Contact: Thomas Genet
- URL: http://people.irisa.fr/Thomas.Genet/timbuk/index.html

3.7. jsexplain

JSExplain

KEYWORDS: JavaScript - Compilation - Standards - Debug - Interpreter

FUNCTIONAL DESCRIPTION: JSExplain is a reference interpreter for JavaScript that closely follows the specification and that produces execution traces. These traces may be interactively investigated in a browser, with an interface that displays not only the code and the state of the interpreter, but also the code and the state of the interpreted program. Conditional breakpoints may be expressed with respect to both the interpreter and the interpreted program. In that respect, JSExplain is a double-debugger for the specification of JavaScript.

- Partner: Imperial College London
- Contact: Alan Schmitt
- Publication: JSExplain: A Double Debugger for JavaScript
- URL: https://gitlab.inria.fr/star-explain/jsexplain

4. New Results

4.1. Compiling Sandboxes:Formally Verified Software Fault Isolation

Participants: Frédéric Besson, Sandrine Blazy, Alexandre Dang, Thomas Jensen.

Software Fault Isolation (SFI) is a security-enhancing pro- gram transformation for instrumenting an untrusted binary module so that it runs inside a dedicated isolated address space, called a sandbox. To ensure that the untrusted module cannot escape its sandbox, existing approaches such as Google's Native Client rely on a binary verifier to check that all memory accesses are within the sandbox. Instead of rely- ing on a posteriori verification, we design, implement and prove correct a program instrumentation phase as part of the formally verified compiler CompCert that enforces a sandboxing security property a priori. This eliminates the need for a binary verifier and, instead, leverages the soundness proof of the compiler to prove the security of the sandbox- ing transformation. The technical contributions are a novel sandboxing transformation that has a well-defined C semantics and which supports arbitrary function pointers, and a formally verified C compiler that im- plements SFI. Experiments show that our formally verified technique is a competitive way of implementing SFI [6].

4.2. Information-Flow Preservation in Compiler Optimisations

Participants: Frédéric Besson, Alexandre Dang, Thomas Jensen.

Correct compilers perform program transformations preserving input/output behaviours of programs. Though mandatory, correctness is not sufficient to prevent program optimisations from introducing information-flow leaks that would make the target program more vulnerable to side-channel at- tacks than the source program. To tackle this problem, we propose a notion of Information-Flow Preserving (IFP) program transformation which ensures that a target program is no more vulnerable to passive side-channel attacks than a source program. To protect against a wide range of attacks, we model an attacker who is granted arbitrary memory accesses for a pre-defined set of observation points. We have proposed a compositional proof principle for proving that a transformation is IFP. Using this principle, we show how a translation validation technique can be used to automatically verify and even close information-flow leaks introduced by standard compiler passes such as dead- store elimination and register allocation. The technique has been experimentally validated on the CompCert C compiler [7].

4.3. Formalization of Higher-Order Process Calculi

Participants: Guillaume Ambal, Alan Schmitt.

Guillaume Amabal and Alan Schmitt, in collaboration with Sergueï Lenglet, have continued exploring how to formalize HOπ in Coq, in particular how to deal with the different kinds of binders used in the calculus. We have extended our previous study that compared locally nameless, De Bruijn indices, and nominal binders with an approach based on higher-order abstract syntax. We have discovered that this approach is not as elegant as in other calculi. A journal version is submitted for publication. The Coq scripts can be found at http://passivation.gforge.inria.fr/hopi/.

4.4. Certified Semantics and Analyses for JavaScript

Participants: Samuel Risbourg, Alan Schmitt.

Alan Schmitt and Samuel Risbourg have continued 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. It was presented to the TC39 committee standardizing JavaScript in December to solicit feedback.

4.5. Skeletal Semantics

Participants: Guillaume Ambal, Nathanael Courant, Thomas Jensen, Adam Khayam, Louis Noizet, Vincent Rebiscoul, Alan Schmitt.

The work on skeletal semantics [5], a modular and formal way to describe semantics or programming languages, has intensified during 2019. We have continued to develop `necro`, a tool to manipulate skeletal semantics and generate interpreters in OCaml, mechanized semantics in Coq, and static analyzers. The code is available online (). Several interns and PhD students are also working on skeletal semantics.

Nathanaël Courant has designed a control-flow analyzer for languages written as skeletal semantics. This work is now extended by Vincent Rebiscoul to certify the analyzer.

Louis Noizet is studying the formalization in Coq of natural semantics from skeletal semantics. To this end, he extended the necro tool to automatically generate a Coq formalization. Louis is also very involved in the maintenance of necro.

Guillaume Ambal is studying the language features that can be captured using skeletal semantics, focusing on concurrency and distribution. In this setting, he is building an approach to automatically derive a small-step semantics form a big-step one.

Adam Khayam is writing a formal semantics of the Hop multitier language, an extension of JavaScript to write web applications. As a first step, he is writing a skeletal semantics of JavaScript to validate that our approach scale for complex and sizable semantics.

4.6. Static analyses for proofs of programs

Participants: Oana Andreescu, Thomas Jensen, Stéphane Lescuyer, Benoît Montagu.

Thomas Jensen together with three industrial research engineers Oana Andreescu, Stéphane Lescuyer, and Benoît Montagu, worked on the development of static analyses that help reduce the manual proof effort that is needed to formally verify programs.

They improved the correlation analysis that Oana Andreescu introduced in her Phd thesis, by designing a novel abstract domain. They verified in Coq its semantic properties, and evaluated their approach on an industrial micro-kernel developed at *Prove&Run*. They showed that the technique could reduce the proof burden by two thirds [1].

4.7. 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 constant-time 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 tool-assisted approaches. This work has been published in [4] as an extended version of [12].

5. Partnerships and Cooperations

5.1. National Initiatives

5.1.1. The ANR Scrypt project

Participants: Frédéric Besson, Sandrine Blazy, Thomas Jensen, David Pichardie, Alexandre Dang, Remi Hutin.

Security, Secure compilation

The Scrypt project (ANR-18-CE25-0014) aims at providing secure implementations of crypto-graphic primitives using formal methods and secure compilation techniques. One specific goal is to design secure compilers which preserve the security of the source code against side-channel attacks.

This is a joint project with the Inria team MARELLE, École Polytechnique and AMOSSYS.

5.1.2. 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 real-time 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.

5.1.3. The ANR AJACS project

Participants: 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 March 2019.

5.1.4. The ANR DISCOVER project

Participants: Sandrine Blazy, David Cachera, Delphine Demange, Thomas Jensen, David Pichardie, Yon Fernandez de Retana, Thomas Rubiano, Yannick Zakowski.

The DISCOVER project project (2014–09/2019) 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 started in October 2014 and ended on September 2019.

5.1.5. The ANR CISC project

Participants: Frédéric Besson, Thomas Jensen, Alan Schmitt.

The goal of the CISC project is to investigate multitier languages and compilers to build secure IoT applications with private communication. In particular, we aim at extending multitier platforms by a new orchestration language that we call Hiphop.js to synchronize internal and external activities of IoT applications as a whole. Our goal is to define language, semantics, attacker models, and policies for the IoT and investigate automatic implementation of privacy and security policies by multitier compilation of IoT applications. To guarantee such applications are correct, and in particular that the required security and privacy properties are achieved, we propose to certify them using the Coq proof assistant. We plan to implement the CISC results as extensions of the multitier language Hop.js (developed at Inria), based on the JavaScript language to maximize its impact. Using the new platform, we will carry out experimental studies on IoT security.

The project partners include the following Inria teams: Celtique, Collège de France, Indes, and Privatics. The project runs from April 2018 to March 2022.

5.2. European Initiatives

5.2.1. FP7 & H2020 Projects

5.2.1.1. The ERC VESTA project

Participants: David Pichardie, Sandrine Blazy, Nicolas Barré, Stefania Dumbrava, Jean-Christophe Léchenet, Rémi Hutin, Aurèle Barrière, Solène Mirliaz.

The VESTA project aims at proposing guidance and tool-support to the designers of static analysis, in order to build advanced but reliable static analysis tools. We focus on analyzing low-level softwares written in C, leveraging on the CompCert verified compiler. Verasco is a verified static analyser that analyses C programs and follows many of the advanced abstract interpretation techniques developped for Astrée. The outcome of the VESTA project will be a platform that help designing other verified advanced abstract interpreters like Verasco, without starting from a white page. We will apply this technique to develop security analyses for C programs. The platform will be open-source and will help the adoption of abstract interpretation techniques.

This a consolidator ERC awarded to David Pichardie for 5 years. The project started in September 2018.

5.2.1.2. The SPARTA cybersecurity competence network

Participants: Thomas Jensen, Frédéric Besson.

SPARTA is a novel Cybersecurity Competence Network, supported by the EU's H2020 program, with the objective to develop and implement top-tier research and innovation collaborative actions. Guided by concrete challenges forming an ambitious Cybersecurity Research & Innovation Roadmap, SPARTA will set up unique collaboration means, leading the way in building transformative capabilities and forming a world-leading Cybersecurity Competence Network across the EU. The SPARTA consortium assembles 44 actors from 14 EU Member States at the intersection of scientific excellence, technological innovation, and societal sciences in cybersecurity.

Celtique is coordinating the Inria participation in the SPARTA network. The team contributes to the programme on intelligent infrastructures with techniques for building security-enhanced systems code that respects strong information flow constraints. The team is also leading the elaboration of the SPARTA scientific roadmap, in collaboration with TU Munich.

5.2.2. 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. 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: Organisation

6.1.1.1. Member of the Conference Program Committees

- Frédéric Besson: PriSC 2019 Workshop on Principles of Secure Compilation
- Sandrine Blazy: POPL 2020 (Symposium on Principles of Programming Languages), APLAS 2019 (Asian Symposium on Programming Languages and Systems), ITP 2019 (Conference on Interactive Theorem Proving), SpiSA 2019 (Workshop on Instruction Set Architecture Specification).
- Thomas Jensen: SAS 2019 (Static Analysis Symposium), ENTROPY 2019 (Enabling Trust through OS proofs).

6.1.1.2. Reviewer

- Frédéric Besson: TACAS

6.1.2. Journal

6.1.2.1. Reviewer - Reviewing Activities

- Frédéric Besson: Journal of Computer Security (JCS), Transactions on Software Engineering (TSE)
- Alan Schmitt: Theoretical Informatics and Applications (RAIRO ITA)

6.1.3. Invited Talks

- Frédéric Besson: GT Sécurité des Systèmes, des Logiciels et des Réseaux
- Sandrine Blazy: *The Verasco static analyzer*, Effective Verification: Static Analysis Meets Program Logics, Lorentz center, Netherlands, May. 2019
- Sandrine Blazy: *Formal Verification of a Constant-Time Preserving C Compiler*, Verified software workshop, Newton institute, Cambridge, United Kingdom, Sept. 2019
- Alan Schmitt: *Sémantiques Formelles et Certifiées*, 30 ans des JFLA
- Alan Schmitt: *Formal JavaScript*, Journées Nationales du GDR GPL
- Alan Schmitt: *Skeletal Semantics*, GT Langages, Types et Preuves

6.1.4. Leadership within the Scientific Community

- Sandrine Blazy is member of Section 6 of the national committee for scientific research CoNRS.
- Sandrine Blazy is member of IFIP WG 2.11 on program generation and of IFIP WG 1.9/2.15 on verified software.

6.1.5. Scientific Expertise

- Sandrine Blazy evaluated two Belgian FWO Grant research proposals.
- Sandrine Blazy is member of the Scientific Advisory Board of the Flemish Cybersecurity Initiative Flanders Program.
- Sandrine Blazy and David Pichardie were members of the HCERES evaluation committee of the LSV laboratory.
- Thomas Jensen was member of the HCERES evaluation committee of the University Grenoble–Alpes LIG laboratory.
- Thomas Jensen was vice-president of the ANR 2019 selection committee for research proposals in Computing and Communication (CES 25).

6.1.6. Research Administration

- Sandrine Blazy and Thomas Jensen are members of the steering committee of the international Static Analysis Symposium.

6.2. Teaching - Supervision - Juries

6.2.1. Teaching

License : Frédéric Besson, Functional programming, 28h, Insa3, Insa Rennes, France

Licence : Sandrine Blazy, Programmation de confiance, 81h, L3, Université Rennes 1, France

Master : Sandrine Blazy, Semantics of Programming Languages, 20h, M1, Université Rennes 1, France

Master : Sandrine Blazy, Software vulnerabilities, 30h, M2, Université Rennes 1, 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 : Thomas Genet, Spécialité Informatique 1 - Algorithmique et Complexité Expérimentale, 47h, L1, Université Rennes 1, France

Licence : Thomas Genet, Initiation au génie logiciel, 67h, L2, Université Rennes 1, France

Licence : Alan Schmitt, Programmation de Confiance, 30h, L3, Université Rennes 1, France

Licence : David Pichardie, Programmation, 30h, L3, ENS Rennes, France

Master : David Pichardie, Semantics of Programming Languages, 30h, M1, Université Rennes 1, France

Master : David Pichardie, Static analysis, 30h, M1, Université Rennes 1, France

Master : Delphine Demange, Software Security, 9h, M2, Université Rennes 1, France

Master : Thomas Genet, Enseigner l'informatique au lycée, 15h, M1, Université Rennes 1, France

Master : Thomas Genet, Analyse et conception formelle, 65h, M1, Université Rennes 1, France

Master : Benoît Montagu, Analyse et Conception Formelles, 16h, M1, Université Rennes 1, France

Master : Alan Schmitt, Suivi de Projet, 24h, M1, ENS Rennes, France

Master : Alan Schmitt, Advanced Semantics, 15h, M2, ENS Rennes, France

Master : Alan Schmitt, Preparation of Agregation exam, 20, M2, ENS Rennes, France

License : Thomas Jensen, Logic, 30h, Insa3, Insa Rennes, France

Master : Thomas Jensen, Software Security , 18h, M2, Université Rennes 1, France

6.2.2. Supervision

Phd : Alexandre Dang, Compilation for memory protection, Université Rennes 1, 10/12/2019, Thomas Jensen, Frédéric Besson

Phd : Julien Lepiller, Verifying Software Fault Isolation, Université Rennes 1, 11/12/2019, Thomas Jensen, Frédéric Besson

PhD in progress : Timothée Haudebourg, Verification of Higher-Order Functional Programs using Tree Automata, September 2017, Thomas Genet and Thomas Jensen

PhD in progress : Rémi Hutin, A C compiler ensuring security properties, September 2018, Sandrine Blazy and David Pichardie

PhD in progress : Aurèle Barrière, Formal verification of a JIT compiler, September 2019, Sandrine Blazy and David Pichardie

PhD in progress : Samy Daoud, Modélisation formelle pour le déploiement et la reconfiguration automatique en ligne dans une infrastructure réseau et IT répartie, November 2018, Alan Schmitt

PhD in progress : Adam Khayam, Formal Semantics of Multitier Languages, July 2019, Alan Schmitt

PhD in progress : Guillaume Ambal, Sémantiques Squelettiques pour Calculs de Processus, September 2019, Alan Schmitt

PhD in progress : Louis Noizet, Compilation Certifiée de Sémantiques Squelettiques, October 2019, Alan Schmitt

Master 2 internship : Aurèle Barrière, Formal verification of a JIT compiler, February of June 2019, Sandrine Blazy and David Pichardie

Master 1 internship : Guillaume Barbier, Maxime Bridoux, Jean Jouve, Clément Legrand-Duchesne, Skeletal Semantics for WebAssembly, September 2018 to May 2019 Alan Schmitt

ENS 4th year internship : Nathanaël Courant, CFA Generation from a Skeletal Semantics, September 2019 to July 2019, Thomas Jensen and Alan Schmitt

ENS 4th year internship : Vincent Rebiscoul, Certified CFA Generation, September 2019 to February 2020, Thomas Jensen and Alan Schmitt

ENS 3th year internship : Justine Sauvage, Modèles formels pour la blockchain, May 2019 to July 2019, Thomas Genet and Thomas Jensen

6.2.3. Juries

- Sandrine Blazy, jury member for the selection of CNRS CR and DR (researchers) candidates, February and March 2019, CNRS, Paris, France.
- Sandrine Blazy, jury member (president) for the PhD defense of Simon Lunel, January 2019, Université Rennes 1
- Sandrine Blazy, jury member for the PhD defense of Joseph Lallemand, November 2019, Lorraine University, Nancy
- Sandrine Blazy, jury member for the HDR defense of Yann Régis-Gianas, November 2019, Paris University
- Sandrine Blazy, jury member (reviewer) for the PhD defense of Mathieu Journault, November 2019, Paris Sorbonne University
- Sandrine Blazy, jury member (president) for the PhD defense of Arif Ali Anapparakkal, December 2019, Université Rennes 1
- Sandrine Blazy, jury member (president) for the PhD defense of Julien Lepiller, December 2019, Université Rennes 1
- Sandrine Blazy, jury member of the GDR GPL PhD award committee
- Sandrine Blazy, member of hiring committee for a professor position, Spring 2019, UBO, Brest
- Thomas Genet, jury member (reviewer) for the PhD defense of Lily Gallois, December 2019, Université de Lille
- Alan Schmitt, president of jury for the PhD defense of Florian Dold, February 2019, Université Rennes 1
- Alan Schmitt, jury member (examiner) for the PhD defense of Julien Lopez, September 2019, Université Paris-Saclay
- Alan Schmitt, jury member (reviewer) for the PhD defense of Steven Varoumas, November 2019, Sorbonne Université
- Delphine Demange, jury member (examiner) for the PhD defense of Oscar Luis Vera Pérez, December 2019, Université Rennes 1
- Delphine Demange, member of hiring committee for a Maître de Conférence position, Spring 2019, Lyon 1 / LIP.

6.3. Popularization

6.3.1. Interventions

- Thomas Genet gave a talk “Bug, Virus, Intrusion, Pirates... So many threats and no defense? Yes... maths.” in high schools close to Rennes. 2019.

7. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] O. ANDREESCU, T. JENSEN, S. LESCUYER, B. MONTAGU.*Inferring frame conditions with static correlation analysis*, in "Proceedings of the ACM on Programming Languages", January 2019, vol. 3, n^{o} POPL, p. 1-29 [*DOI :* 10.1145/3290360], https://hal.inria.fr/hal-02413262

[2] 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", April 2019, vol. 62, n^{o} 4, p. 433-480 [*DOI :* 10.1007/S10817-017-9439-Z], https://hal.inria.fr/hal-01656895

[3] F. BESSON, S. BLAZY, P. WILKE.*CompCertS: A Memory-Aware Verified C Compiler using a Pointer as Integer Semantics*, in "Journal of Automated Reasoning", August 2019, vol. 63, n^{o} 2, p. 369-392 [*DOI :* 10.1007/S10817-018-9496-Y], https://hal.inria.fr/hal-02401182

[4] S. BLAZY, D. PICHARDIE, A. TRIEU.*Verifying constant-time implementations by abstract interpretation*, in "Journal of Computer Security", 2019, vol. 27, n^{o} 1, p. 137–163 [*DOI :* 10.3233/JCS-181136], https://hal.inria.fr/hal-02025047

[5] M. BODIN, P. GARDNER, T. JENSEN, A. SCHMITT.*Skeletal Semantics and their Interpretations*, in "Proceedings of the ACM on Programming Languages", 2019, vol. 44, p. 1-31, forthcoming [*DOI :* 10.1145/3290357], https://hal.inria.fr/hal-01881863

International Conferences with Proceedings

[6] F. BESSON, S. BLAZY, A. DANG, T. JENSEN, P. WILKE.*Compiling Sandboxes: Formally Verified Software Fault Isolation*, in "ESOP 2019 - 28th European Symposium on Programming", Prague, Czech Republic, LNCS, Springer, April 2019, vol. 11423, p. 499-524 [*DOI :* 10.1007/978-3-030-17184-1_18], https://hal.inria.fr/hal-02316189

[7] F. BESSON, A. DANG, T. JENSEN.*Information-Flow Preservation in Compiler Optimisations*, in "CSF 2019 - 32nd IEEE Computer Security Foundations Symposium", Hoboken, United States, IEEE, June 2019, p. 1-13, https://hal.inria.fr/hal-02180303

[8] S. BLAZY.*Teaching Deductive Verification in Why3 to Undergraduate Students*, in "FM Tea (Formal Methods Teaching)", Porto, Portugal, September 2019, p. 52-66 [*DOI :* 10.1007/978-3-030-32441-4_4], https://hal.inria.fr/hal-02362306

[9] S. BLAZY, R. HUTIN.*Formal Verification of a Program Obfuscation Based on Mixed Boolean-Arithmetic Expressions*, in "CPP 2019 - 8th ACM SIGPLAN International Conference on Certified Programs and Proofs", Cascais, Portugal, ACM, January 2019, p. 196-208 [*DOI :* 10.1145/3293880.3294103], https://hal.inria.fr/hal-01955773

Conferences without Proceedings

[10] P. TALBOT, D. CACHERA, E. MONFROY, C. TRUCHET.*Octogones entiers pour le problème RCPSP*, in "JFPC 2019 - Journées Francophones de Programmation par Contraintes", Albi, France, June 2019, p. 1-10, https://hal.archives-ouvertes.fr/hal-02157804

Research Reports

[11] S. MIRLIAZ, D. PICHARDIE.*Flow insensitive relational static analysis*, ENS Rennes ; Université Rennes 1, June 2019, https://hal.archives-ouvertes.fr/hal-02332139

References in notes

[12] 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

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

Project-Team CIDRE

Creation of the Project-Team: 2011 July 01

Keywords:

Computer Science and Digital Science:

A1.1.8. - Security of architectures
A1.2.3. - Routing
A1.2.8. - Network security
A1.3. - Distributed Systems
A1.3.3. - Blockchain
A1.3.4. - Peer to peer
A1.3.5. - Cloud
A2.3.1. - Embedded systems
A3.1.5. - Control access, privacy
A3.3.1. - On-line analytical processing
A3.4.1. - Supervised learning
A3.4.2. - Unsupervised learning
A3.5.2. - Recommendation systems
A4.1. - Threat analysis
A4.1.1. - Malware analysis
A4.1.2. - Hardware attacks
A4.4. - Security of equipment and software
A4.5. - Formal methods for security
A4.8. - Privacy-enhancing technologies
A4.9.1. - Intrusion detection
A4.9.2. - Alert correlation

Other Research Topics and Application Domains:

B6.3.3. - Network Management
B6.5. - Information systems
B9.6.2. - Juridical science
B9.10. - Privacy

1. Team, Visitors, External Collaborators

Research Scientists

Ludovic Mé [Inria, Advanced Research position, HDR]
Emmanuelle Anceaume [CNRS, Researcher, HDR]
Michel Hurfin [Inria, Researcher, HDR]
Jean-Louis Lanet [Inria, Researcher, HDR]

Faculty Members

Christophe Bidan [Centrale-Supélec, Professor, HDR]
Valérie Viet Triem Tong [Team leader, Centrale-Supélec, Associate Professor, HDR]
Gilles Guette [Univ de Rennes I, Associate Professor]

Guillaume Hiet [Centrale-Supélec, Associate Professor]
Jean-Francois Lalande [Centrale-Supélec, Professor, HDR]
Guillaume Piolle [Centrale-Supélec, Associate Professor]
Sam Thomas [Centrale-Supélec, Associate Professor, from Jun 2019 until Nov 2019]
Eric Totel [Centrale-Supélec, Professor, until Sep 2019, HDR]
Frédéric Tronel [Centrale-Supélec, Associate Professor]
Pierre Wilke [Centrale-Supélec, Associate Professor]

External Collaborator

Frédéric Majorczyk [DGA]

Technical Staff

Mohamed Alsamman [Inria, Engineer, from Sep 2019]

PhD Students

Nicolas Bellec [Inria, from Oct 2019]
Aimad Berady [from Nov 2018]
Sebanjila Bukasa [Inria, until Mar 2019]
Vasile Cazacu [CNRS]
Ronny Chevalier [HP France]
Tomas Javier Concepcion Miranda [Centrale-Supélec, from Oct 2019]
Alexandre Dey [Airbus]
Aurélien Dupin [Thales, until Jan 2019]
Mathieu Escouteloup [Inria]
Benoit Fournier [Univ de Rennes I]
Cyprien Gottstein [Orange Labs]
Pierre Graux [Inria]
Cedric Herzog [Inria]
David Lanoé [Inria, until Sep 2019]
Kevin Le Bon [Inria]
Laetitia Leichtnam [Ministère de la Défense, until Sep 2019]
Leopold Ouairy [Inria]
Aurelien Palisse [Inria, until Jan 2019]
Charles Arya Xosanavongsa [Thales, until Nov 2019]

Post-Doctoral Fellows

Ludovic Claudepierre [Inria until Oct 2019,Univ de Rennes I since Nov 2019]
Jerome Fellus [Univ de Rennes I, until Jun 2019]
Mouad Lemoudden [Inria]
Frédérique Robin [Inria, from Oct 2019]

Administrative Assistants

Lydie Mabil [Inria]
Alexandre Dang [Centrale-Supélec, from Oct 2019]

2. Overall Objectives

2.1. CIDRE in Brief

The Cidre team is concerned with security and privacy issues. Our long-term ambition is to contribute to the construction of widely used systems that are trustworthy and respectful of privacy, even when parts of the system are targeted by attackers.

With this objective in mind, the CIDRE team focuses mainly on the three following topics:

- **Attack comprehension**
- **Attack detection**
- **Attack resistance**

3. Research Program

3.1. Our perspective

For many aspects of our daily lives, we rely heavily on computer systems, many of which are based on massively interconnected devices that support a population of interacting and cooperating entities. As these systems become more open and complex, accidental and intentional failures become much more frequent and serious. We believe that the purpose of attacks against these systems is expressed at a high level (compromise of sensitive data, unavailability of services). However, these attacks are often carried out at a very low level (exploitation of vulnerabilities by malicious code, hardware attacks).

The CIDRE team is specialized in the defense of computer systems. We argue that to properly protect these systems we must have a complete understanding of the attacker's concrete capabilities. In other words, **to defend properly we must understand the attack**.

The CIDRE team therefore strives to have a global expertise in information systems: from hardware to distributed architectures. Our objective is to highlight security issues and propose preventive or reactive countermeasures in widely used and privacy-friendly systems.

3.2. Attack Comprehension

An attack on a computer system begins with the exploitation of one or more vulnerabilities of that system. Generally speaking, a vulnerability can be a software bug or a misconfiguration that can be exploited by the attacker to perform unauthorized actions. Exploiting a vulnerability leads to a use of the system according to a case not foreseen in its specification, implementation or configuration. This puts the system in an inconsistent state allowing the attacker to divert the use of the system in his or her own interest.

The systems we use are large, interconnected, constantly evolving and, therefore, are likely to retain many vulnerabilities; their security depends on our ability to update them quickly when new threats are discovered. It is thus necessary to understand how the attacker has compromised the system: what vulnerabilities he has exploited, what actions he has conducted, where he is located in the system. It is essential to study statically the malicious code used by the attacker. It is also important to be able to study it dynamically to be able to replay attacks on demand.

Ideally, we should be ahead of the attacker and therefore imagine new ways to attack. In addition, we believe it is necessary to improve the feedback to the expert by allowing him to quickly understand the progress of an attack. The first step before being able to offer secure systems is to understand and measure the real capabilities of the attacker.

Our first research axis therefore aims at highlighting both the effective attacker's means and the way an attack unfolds and spreads.

In this context, we are particularly interested in

- **highlighting attacks** on the micro-architecture that affect software security
- **providing expert support**
  - to analyze malicious code
  - to quickly investigate an intrusion on a system monitored by an intrusion detection system

3.3. Attack Detection

An attack is generally composed of several steps. During a first approach step the attacker enters the system, locates the target and makes itself persistent. Then, in a second step, the payload of the attack is effectively launched, leading to a violation of the security policy (attacks against confidentiality, integrity, or availability of OS, applications, services, or data).

The objective of intrusion detection is to be able to detect the attacker, ideally during the first step of the attack. To do this, intrusion detection systems (IDS) are based on probes that continuously monitor the system. These probes report events to a core engine that decide whether or not to alert the expert.

Intrusion detection systems are important for all systems handling sensitive data that may be accessible to a malicious agent. They are especially crucial for low-level systems that provide essential support services to other systems. They are essential in inter-connected systems that are designed to last a long time and are difficult to update.

3.4. Attack Resistance

The first two axes of the team allowed us to measure the concrete technical means of the attacker. We claim that the attacker can always avoid the measures put in place to secure a system. We believe that another way to offer more secure systems is to take into account from the design phase that these systems will operate in the presence of an omnipotent attacker. The last research axis of the CIDRE team is focused on offering systems that are resistant to attackers, *i.e.* they can provide the expected services even in the presence of an attacker.

To achieve this goal, we explore two approaches:

- deceptive security
- malicious behavior tolerance

In the notion of *deceptive security* we group together all the approaches that aim to mislead the active attacker in a system in order to deceive him on the exact nature of his target. These approaches can slow down the attacker or lead him to abandon his attack.

Finally, we contribute to the design of architectures or services relying on the collaboration of entities that is not affected by the minority presence of malicious entities. These architectures or services are based on the collaboration of a set of nodes that are not affected by the presence in minority of malicious nodes.

4. Highlights of the Year

4.1. Highlights of the Year

This year we highlight two key events in the team's life:

- We have organized the SILM semester on the Security of Software/Hardware Interfaces. The goal of this semester is to promote the scientific, teaching and industrial transfer activities on the security of software/hardware interfaces. This semester is supported by DGA.
- We have concluded the transfer of a license to use GroddDroid our Android malware analysis framework.

5. New Software and Platforms

5.1. 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

5.2. 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.

- Authors: Mourad Leslous, Adrien Abraham, Pierre Graux, Jean François Lalande, Valérie Viet Triem Tong and Pierre Wilke
- 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

5.3. 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

5.4. GroddViewer

KEYWORDS: Android - Detection - Malware

FUNCTIONAL DESCRIPTION: To visualise data from GroddDroid

- Authors: Jean-François Lalande, Valérie Viet Triem Tong, Sébastien Campion, Mathieu Simon and Pierre Wilke
- Contact: Valérie Viet Triem Tong

5.5. Survivor

KEYWORDS: Intrusion Response - Intrusion Recovery - Survivability - Resiliency - Linux - Checkpoint/Restore - Threat Mitigation

FUNCTIONAL DESCRIPTION: Survivor is a set of low-level components to design a Linux-based operating system able to withstand ongoing intrusions and to allow business continuity despite the presence of an active adversary. Survivor provides an Intrusion Response System (IRS) with the low-level components and interfaces needed to orchestrate a per-service checkpoint, recovery, and mitigation actions. It recovers infected services (i.e., their processes and their associated files) to a previous safe state and it protects their state by applying a set of mitigations (e.g., privilege restrictions and resource quotas) aimed at withstanding further reinfections.

- Participants: Ronny Chevalier, Guillaume Hiet, David Plaquin and Chris Dalton
- Partners: CentraleSupélec - HP Labs
- Contact: Ronny Chevalier

5.6. PyMaO

Python Malware Orchestrator

KEYWORDS: Android - Malware

FUNCTIONAL DESCRIPTION: PyMaO chains several analyses that are part of an experiment. An analysis is most of the time, a call to an external tool that returns a result, for example apktool, grep, Androguard, Apkid. An experiment is a collection of analyses that are run one by one, chained, if some conditions hold. For example, if the unpacking of an application with Apktool succeeds, then you can grep the code for searching a string.

PyMaO has a nice old-fashion graphical interface (ncurses).

RELEASE FUNCTIONAL DESCRIPTION: Initial release corresponding to the demo presented at MASCOTS 2019.

NEWS OF THE YEAR: A demo has been presented at the MASCOTS 2019 conference: https://hal-centralesupelec.archives-ouvertes.fr/hal-02305473

- Authors: Jean-François Lalande, Pierre Graux and Tomas Javier Concepcion Miranda
- Contact: Jean-François Lalande
- URL: https://gitlab.inria.fr/cidre-public/pymao

5.7. OATs'inside

KEYWORDS: Android - Malware - Reverse engineering - Code analysis

FUNCTIONAL DESCRIPTION: OATs'inside is a Android reverse engineering tool that handles all native obfuscation techniques. This tool uses a hybrid approach based on dynamic monitoring and trace-based symbolic execution to output control flow graphs (CFGs) for each method of the analyzed application. These CFGs spare users the need to dive into low-level instructions, which are difficult to reverse engineer.

- Participants: Pierre Graux, Jean-François Lalande, Valérie Viet Triem Tong and Pierre Wilke
- Contact: Pierre Graux

6. New Results

6.1. Axis 1 : Attack comprehension

6.1.1. Fault injection

Electromagnetic injection is a non-invasive way to attack a chip. The large number of parameters that require to be properly tuned for such an attack limits its efficiency. In [30] we propose several ways to improve the success rate of fault injection by electromagnetic radiation. We show that software execution is altered at targeted instructions if the radiating probe is located above the phase-locked loop device driving the clock tree. We identify the phase-locked loop as a sensitive part of the chip. We reduce the preferential location for the electromagnetic injection to a small area in the vicinity of the analog power supply feeding the phase-locked loop. We also explore the influence of the frequency of the injected electromagnetic wave. We compute the optimal fault rate in a bandwidth of $15MHz$, in the upper limit of the chip bandwidth. Our experiments show that for an optimal frequency a precision of $5ns$, we succeed to reach the best fault rate. With this electromagnetic injection technique, the achieved success rate reaches 15 to 20%. Such a fault can be used to retrieve the key of a cryptographic algorithm (for an Advanced Encryption Standard application for example).

6.1.2. Malware analysis

About Android malware analysis, we have started investigations with specific malware that hide their behavior using obfuscation techniques [10]. As these malware are difficult to find in the wild, we have also started to analyze both datasets of the literature and large collection of applications captured from different repositories such as the Play Store. This huge amount of applications to analyze (currently more than 100,000) makes difficult to build reliable experiments [20]. We have designed a new tool, called PyMaO, that helps to orchestrate experiments. This tool is published as an open source tool under GPL v3. We have also revisited the historical datasets of malware of the literature and introduce a more up-to-date malware and goodware dataset [26].

6.1.3. Focus on doxware

A doxware is a particular type of ransomware that threatens to release personal or sensitive data to the public if the user does not pay the ransom. The term comes from the hacker term "doxing," or releasing confidential information over the internet. The only difference between a classical ransomware and a doxware resides in a *valuable files hunting* followed by an exfiltration of these data. In [34], we have explored how an attacker may be able to quickly localized valuable assets of a machine using an analysis of the content and the vocabulary of its files.

6.1.4. Attack scenario reconstruction

In order to supervise the security of a large infrastructure, the administrator deploys multiple sensors and intrusion detection systems on several critical places in the system. It is easier to explain and detect attacks if more events are logged. Starting from a suspicious event (appearing as a log entry), the administrator can start his investigation by manually building the set of previous events that are linked to this event of interest. Accordingly, the administrator attempts to identify links among the logged events in order to retrieve those that correspond to the traces of the attacker's actions in the supervised system; previous work is aimed at building these connections. In practice, however, this type of link is not trivial to define and discover. Hence, there is a real necessity to describe and define formally the semantics of these links in literature. In In order to supervise the security of a large infrastructure, the administrator deploys multiple sensors and intrusion detection systems on several critical places in the system. It is easier to explain and detect attacks if more events are logged. Starting from a suspicious event (appearing as a log entry), the administrator can start his investigation by manually building the set of previous events that are linked to this event of interest. Accordingly, the administrator attempts to identify links among the logged events in order to retrieve those that correspond to the traces of the attacker's actions in the supervised system; previous work is aimed at building these connections. In practice, however, this type of link is not trivial to define and discover. Hence, there is a real necessity to describe and define formally the semantics of these links in literature. In this paper, a clear definition of this relationship, called contextual event causal dependency, is introduced and proposed. The work presented in this paper aims at defining a formal model that would ideally unify previous work on causal dependencies among heterogeneous events. We define a relationship among events that enables the discovery of all events, which can be considered as the cause (in the past) or the effect (in the future) of an event of interest (e.g., an indicator of compromise, produced by an attacker action). In [36], we have proposed a clear definition of this relationship, called contextual event causal dependency. The work presented in [36] aims at defining a formal model that would ideally unify previous work on causal dependencies among heterogeneous events. We define a relationship among events that enables the discovery of all events, which can be considered as the cause (in the past) or the effect (in the future) of an event of interest (e.g., an indicator of compromise, produced by an attacker action).

6.2. Axis 2 : Attack detection

6.2.1. Vulnerabilities detection in Java

In a prior work, we have focused on adapting a machine-learning tool (ChuckyJava) aiming at automatically detect vulnerabilities in Java. ChuckyJava is able to detect vulnerabilities by performing in two steps: the neighborhood discovery and the anomaly detection. The neighborhood discovery is the ability for the tool to detect method of similar semantics: neighbors. In [25], we mitigate many ChuckyJava's limitations by developing JavaNeighbors that improves the neighborhood discovery. JavaNeighbors represents methods by terms and using a method based on a Natural Language Processing technique, JavaNeighbors computes the distance between all representations of methods. Finally, according to the distance, each method has a neighbor list from the closest to the most distant ones. JavaNeighbors has enabled ChuckyJava to detect vulnerabilities with more accuracy.

6.2.2. Ransomware detection

A ransomware attacks mostly begins with social engineering methods to install payloads on victims' computers, followed by a communication with command and control servers for data exchange. To enable an early detection and thus scale down these attacks, we propose in [35] a detection model based on the collected system and network logs from a computer. The analysis is performed on various ransomware families with a high detection rate. Packet level detection is performed to grant the best use case scenario. This work intends to provide an independent third-party procedure that is able to distinguish between a benign software and a malicious ransomware based on network activity. Furthermore, it is not limited to only identify ransomware but could be utilized to inspect different malware.

6.2.3. Intrusion detection using logs of distributed application

Although security issues are now addressed during the development process of distributed applications, an attack may still affect the provided services or allow access to confidential data. To detect intrusions [22], we consider an anomaly detection mechanism which relies on a model of the monitored application's normal behavior. During a model construction phase, the application is run multiple times to observe some of its correct behaviors. Each gathered trace enables the identification of significant events and their causality relationships, without requiring the existence of a global clock. The constructed model is dual: an automaton plus a list of likely invariants. The redundancy between the two sub-models decreases when generalization techniques are applied on the automaton. Solutions already proposed suffer from scalability issues. In particular, the time needed to build the model is important and its size impacts the duration of the detection phase. The proposed solutions address these problems, while keeping a good accuracy during the detection phase, in terms of false positive and false negative rates. To evaluate them, a real distributed application and several attacks against the service have been considered. One of our goal is to identify redundancies and complementarities between the proposed models.

6.3. Axis 3 : Attack resistance

6.3.1. Attacker Life cycle

We have been witnessing for years the awareness of the existence of a so-called Advanced Persistent Threat (APT). These attacks, regularly target or involving nation-states and large companies, were first defined in 2011. Ad Advanced Persistent Threat: (i) pursues its objectives repeatedly over an extended period of time; (ii) adapts to defenders' efforts to resist it; and (iii) is determined to maintain the level of interaction needed to execute its objectives. In [13], we have proposed a model providing an operational reading of the attackers' lifecycle in a compromised network. This model allows to express possible regressions in the attack and introduces the concept of a waiting state, which is essential for long-term actions. In this article we have also proposed a confrontation between our model and two recent examples of attacks whose progression has been publicly described: the Equifax breach (2017) and the TV5Monde sabotage (2015).

6.3.2. OS-level intrusion survivability

Despite the deployment of preventive security mechanisms to protect the assets and computing platforms of users, intrusions eventually occur. In [17], we have proposed a novel intrusion survivability approach to withstand ongoing intrusions. Our approach relies on an orchestration of fine-grained recovery and per-service responses (e.g., privileges removal). Such an approach may put the system into a degraded mode. This degraded mode prevents attackers to reinfect the system or to achieve their goals if they managed to reinfect it. It maintains the availability of core functions while waiting for patches to be deployed. We devised a cost-sensitive response selection process to ensure that while the service is in a degraded mode, its core functions are still operating. We built a Linux-based prototype and evaluated the effectiveness of our approach against different types of intrusions. The results show that our solution removes the effects of the intrusions, that it can select appropriate responses, and that it allows services to survive when reinfected. In terms of performance overhead, in most cases, we observed a small overhead, except in the rare case of services that write many small files asynchronously in a burst, where we observed a higher but acceptable overhead.

6.3.3. Secure routing in drones swarms

Unmanned aerial vehicle (UAV) applications and development have increased over the past few years as this technology has become more accessible and less expensive. On a single UAV scenario, communication is a keystone to transmit commands and retrieve data from UAV sensors. It is even more critical in swarm where cooperation and inter messaging is fundamental. The communication between the nodes of a swarm is based on a suitable routing algorithm. The routing must allow each node to send messages to each other, by successive hops between different neighbors. A UAV swarm is a particular mobile ad-hoc networks where nodes run independently but form a cooperative communication network. UAV swarm shares com- mon characteristics with VANET (vehicular ad hoc network), sensors network or mobile phone network but also strongly differs on specific points (mobility model, instability, limited infrastructure access). Any computation on a UAV is a permanent trade off between volume, weight and power consumption, with no infrastructure access. In [31], we have proposed a secured routing protocol designed for UAV swarm networks. SEER4US is the first protocol providing integrity of routing messages and authentication of their sender with low energy consumption for battery preservation.

6.3.4. Securing the control flow of smartcard C programs

Results obtained several years ago about securing the control flow of C programs have been extended and published in the journal Computers and Security [7]. This extended version of our work focuses on the formal verification of the introduced countermeasures. We prove that any possible attack that would skip more than one C instruction is detected by our countermeasures. We also extended the experimental results on a benchmark software dedicated to smartcards. This work has been achieved in cooperation with Karine Heydemann from the LIP6 laboratory (Sorbonne Université).

6.3.5. A secure implementation of the replicated state machine

State machine replication (RSM) is today the foundation of many cloud-based highly-available products: it allows some service to be deployed such to guarantee its correct functioning despite possible faults. In RSM, clients issue operation requests to a set of distributed processes implementing the replicated service, that, in turn, run a protocol to decide the order of execution of incoming operations and provide clients with outputs. Faults can be accidental (e.g. a computer crashing due to a loss of power) or have a malicious intent (e.g. a compromised server). Whichever is the chosen fault model, RSM has proven to be a reliable and effective solution for the deployment of dependable services. RSM is usually built on top of a distributed Consensus primitive that is used by processes to agree on the order of execution of requests concurrently issued by clients. The main problem with this approach is that Consensus is impossible to achieve deterministically in a distributed settings if the system is asynchronous and even just a single process may fail by crashing. This led the research community to study and develop alternative solutions based on the relaxation of some of the constraints, to allow agreement to be reached in partially synchronous systems with faulty processes by trading off consistency with availability. An alternative approach consists in imposing constraints on the set of operations that can be issued by clients, i.e. imposing updates that commute. In particular, commutative replicated data types (CRDTs) can be implemented with an RSM approach in asynchronous settings using the monotonic growth of a join semilattice, i.e., a partially ordered set that defines a join (least upper bound) for all element pairs. In [18] we have proposed an algorithm that solves Generalized Lattice Agreement in a Byzantine fault model. To the best of our knowledge this is the first solution for Byzantine lattice agreement that works on any possible lattice, and it is the first work proposing a Byzantine tolerant RSM built on it. The algorithm is wait-free, i.e., every process completes its execution of the algorithm within a bounded number of steps, regardless of the execution of other processes. We have also sketch the main lines of a signature-based version of our algorithms which take advantage of digital signatures to reduce the message complexity to $\mathcal{O}(n)$ per process, when the number f of Byzantine processes verifies $f = \mathcal{O}(1)$.

6.3.6. Blockchain in adversarial environments

We are pursuing our efforts dedicated to the theoretical aspects of blockchains. In particular, we have recently proposed to specify blockchains as a composition of abstract data types all together with a hierarchy of consistency criteria that formally characterizes the histories admissible for distributed programs that use them. Our work is based on an original oracle-based construction that, along with new consistency definitions, captures the eventual convergence process in blockchain systems. This study allows us to focus on the implementability of the presented abstractions and a mapping of representative existing blockchains from both academia and industry in our framework. It is already known that some blockchain implementations solve eventual consistency of an append-only queue using Consensus. However the question about the consistency criterion of blockchains as Bitcoin and Ethereum that technically do not solve Consensus, and their relation with Consensus in general was not studied. We have also proposed a specification of distributed ledger register that matches the Lamport hierarchy from safe to atomic. Moreover, we propose implementations of distributed ledger registers with safe, regular and atomic guaranties in a model of communication specific to distributed ledgers technology that we also formalize. Then, we propose an implementation of a distributed ledger register that satisfies the atomic specification and the k-consistency property that characterizes the permissionless distributed blockchains such as Bitcoin and Ethereum. Preliminary results appear in [41].

In parallel to this work, we have proposed the design of a scalable permissionless blockchain in the proof-of-stake setting. In particular, we use a distributed hash table as a building block to set up randomized shards, and then leverage the sharded architecture to validate blocks in an efficient manner. We combine verifiable Byzantine agreements run by shards of stakeholders and a block validation protocol to guarantee that forks occur with negligible probability. We impose induced churn to make shards robust to eclipse attacks, and we rely on the UTXO coin model to guarantee that any stake-holder action is securely verifiable by anyone. Our protocol works against adaptive adversary, and makes no synchrony assumption beyond what is required for the byzantine agreement. This work has been published in [19].

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

- **HP (2013-2019): Embedded Systems Security** One of the main activities of HP Inc. is to develop and manufacture computing platforms (such as laptops, printers, etc). These platforms consist of hardware and embedded software (usually referred to as firmware). Such embedded software is typically required for the proper functioning of the hardware and relied upon by high level operating system, application or solution software. One of the research tracks of this collaboration consists in enhancing the security level of low-level software components (firmware and OS) in future computing platforms. The final objective is to provide a more resilient and trustworthy platform to the end-user. This work is carried out in the context of the PhD of Ronny Chevalier.
- **DGA (2018-2020)** Traditionally, IDSes are evaluated based on their detection ability against a labeled dataset that contains normal and abnormal network traffic. Upon inspection, it is clear that datasets publicly available are usually obsolete in the span of a couple years in both anomaly types and background, benign Internet traffic. They also suffer from a lack of volume and diversity in traffic, and ultimately, lack of representativeness and realism. In this context, the goal of this project is to come up with an evolutive platform for IDS evaluation that solves many of the issues that exist in the state of the art methods. In order to create such an evolutive platform, there is a need for dynamic infrastructure that allows continuous and automatic change. Here are a number of design principles that we followed for our platform: reproducibility (it is possible to rebuild the infrastructure of the platform or any element of it); repeatability (any action carried out on the infrastructure tested in the platform is repeatable); live evaluation (while traditional IDS evaluation is carried out using a static benchmark dataset, we propose an environment that resembles what IDS does in real life); realism (in terms of traffic generation, real world attack representativeness, and system setup. This will surely be a continuous and evolutive effort to try to approach real world conditions as best as can be); automatization (scripts allow a complete description of the system in which an IDS is tested, and of normal/malicious activity generation inside this system).

  This work is carried out in the context of the postdoc of Mouad Lemoudden.
- **DGA (2019-2021)** DGA and its industrial partners have to regularly implement filters applied to standard or proprietary protocols on communication interfaces or directly in products. In order to allow administrators to easily adapt these filters to the specific context of the various devices, filtering languages specific to the different filtering policies applicable to the different devices should be developed. Even for simple static filters, the definition of such languages is a complex task. A methodological approach that would simplify this task for higher level abstraction filtering languages (and therefore simpler to use) would be to allow the definition of higher level abstraction filtering languages by relying on a single language of lower level of abstraction. This would make it possible to define high-level abstraction and easy-to-use languages in a recursive way by progressively

increasing the levels of abstraction (and specificity). In addition, this approach would improve reusability. Indeed, it would be possible to rely on a filtering language, previously developed for another project, in order to more easily develop a more specific (and easy to use) language for another project.

This work is carried out in the context of the postdoc of Ludovic Claudepierre

7.2. Bilateral Grants with Industry

- **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.
- **Idemia: Hardware Security for Embeded Devices** Kevin Bukasa has started his PhD in January 2016 in a bilateral contract between Inria and Idemia. He explored fault injection attacks using EM probes on two different kind of devices: microcontroller (representing IoT) and SoC (representing Smart phone). He demonstrated the vulnerability of both architectures on this kind of attack. On IoT device he has developped an attack allowing to take a full control on the device. He discovered also new fault attacks never described in the litterature.
- **Idemia: Protection against fuzzing attack** Leopold Ouairy has started his PhD in October 2017 in a bilateral contract between Inria and Idemia. The context is related with security testing of Java applications to avoid fuzzing attack. The approach is based on AI to design automatically a model use for the oracle. He used machine learning to serach in a corpus of applicatons methods having the same semantics. Then in a second step, after convertir the source code into a vector he compute a similarity value which is related with absence of conditions evaluation.
- **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.
- **Ministry of Defence: Characterization of an attacker** Aïmad Berady has started his PhD thesis in November 2018 in the context of a contract between CentraleSupelec and the French Ministry of Defence. His work is to highlight the characteristics of an attacker performing a targeted and long-term attack on an information system.
- **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 multi-tenant 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.
- **Orange LAb's: Storage and query in a massive distributed graph for the web of things** Cyprien Gottstein has started his PhD thesis in October 2018 in the context of a collaboration between Inria and Orange (I/O Lab). In this thesis, we consider storage and query problems that arise when massive distributed graphs are used to represent the web of things. In particular, access to the data and partitioning of the graph are studied to propose efficient geographical services.

- **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.

8. Partnerships and Cooperations

8.1. Regional Initiatives

- **Labex COMINLABS contract (2016-2019): "BigClin" - https://bigclin.cominlabs.u-bretagneloire.fr/fr**

  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.

8.2. National Initiatives

- **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.

8.3. International Research Visitors

8.3.1. Research Stays Abroad

Emmanuelle Anceaume has been invited by the University of La Sapienza (Italy) from the 1st to the 30th of September 2019. During this stay, she collaborated with Profs Leonardo Querzony and Giuseppe A. Di Luna. Their collaboration gave rise to an implementation of the Replicated State Machine, which is resilient to Byzantine behaviors in asynchronous environments [18] (will appear at IPDPS in 2020).

8.4. European Initiatives

8.4.1. H2020 Projects

- **SPARTA (2019-2022) - https://www.sparta.eu/**

  SPARTA is a Cybersecurity Competence Network supported by the EU's H2020 program (Grant agreement ID: 830892) and led by CEA. This 3 years project started in February 2019. It aims to coordinate and develop the implementation of high-level research and innovation in digital security, in order to strengthen the strategic autonomy of the European Union.The CIDRE team is involved both in the workpackage 2 (SPARTA Roadmap) that aims to develop an ambitious Cybersecurity Research and Innovation Roadmap and the workpackage 6 (SPARTA Program HAII-T) that will develop a foundation for secure-by-design Intelligent infrastructures. More precisely, in the context of a task dedicated to resilience-by-design, we design an intrusion detection mechanism that combines both signature-based and anomaly-based approaches.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Organisation

9.1.1.1. General Chair, Scientific Chair

- Ludovic Mé served as general chair of the conference RESSI (Rendez-Vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information), May 2019, Erquy, France.
- Guillaume Hiet is the chair holder of the SILM thematic semester. He served as general chair of the SILM workshop, November 2019, Rennes, France.

The SILM thematic semester [0] is dedicated to the security of software/hardware interfaces and focus more particularly on the three following axes:

[0] https://silm.inria.fr/

1. Analyzing the behavior and the state of hardware components using, e.g. trace mechanisms, fuzzing, reverse-engineering techniques, or side channel analyses;
2. Studying the hardware vulnerabilities and the software attacks that can exploit them: e.g. side-channels, fault injections, or exploitation of unspecified behavior;
3. Detecting and preventing software attacks using dedicated hardware components. Proposing software countermeasures to protect from hardware vulnerabilities.

The goal of this semester is to promote the scientific, teaching and industrial transfer activities on the security of software/hardware interfaces. Our objective is also to identify scientific and technological challenges in that field and to propose a strategic action plan. To that end, we organized different events:

- The SILM summer school (in collaboration with the GDR "Sécurité Informatique") [0], in July 2019, Rennes, France.
- The SILM workshop [0], in Novembre 2019, Rennes, France
- A regular seminar [0] at Inria, Rennes, France.

We will also animate a working group and publish a white-paper on that topic.

9.1.1.2. Member of the Organizing Committees

Christophe Bidan served as a member of the organization committee of C&ESAR 2019 (26rd Computers & Electronics Security Applications Rendez-vous), November 2019, Rennes, France.

Eric Totel served as a member of the organization committee of the conference RESSI (Rendez-Vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information), May 2019, Erquy, France.

Gilles Guette served as a member of the organization committee of the conference RESSI (Rendez-Vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information), May 2019, Erquy, France.

Frédéric Tronel served as a member of the organization committee of SSTIC 2019 (Symposium sur la sécurité des technologies de l'information et des communications) that took place in Rennes, France in June, where it gathered more than 600 participants.

Frédéric served as a member of the organisation committee of SILM Summer School and SILM workshop.

9.1.2. Scientific Events: Selection

9.1.2.1. Chair of Conference Program Committees

Jean-François Lalande was the program co-chair of HPCS 2019 (The 2019 International Conference on High Performance Computing & Simulation), July 2019, Dublin, Ireland, IEEE Computer Society.

Jean-François Lalande was the program co-chair of CECC 2019 (Central European Cybersecurity Conference), November 2019, Munich, Germany, ACM Press.

Jean-François Lalande was the program co-chair of IWSMR 2019 workshop (1st International Workshop on Information Security Methodology and Replication Studies), August 2019, Dublin, Ireland, ACM Press.

9.1.2.2. Member of the Conference Program Committees

Ludovic Mé served the Scientific Committee of the FIC 2020.

Frédéric Tronel and Valérie Viet Triem Tong served as a member of the program committee of SSTIC 2019 (Symposium sur la sécurité des technologies de l'information et des communications) June 2019, Rennes, France.

[0] https://silm-school.inria.fr/
[0] https://silm-workshop.inria.fr/
[0] https://semestres-cyber.inria.fr/en/silm-seminar/

Jean-François Lalande served as a member of the program committee of the international conferences CECC 2019, SecITC 2019, and of international workshops IWCC 2019, CUING 2019, SHPCS 2019, WTMC 2019.

Gilles Guette served as a member of the program committee of the international conferences ICISSP 2019, ISNCC 2019, ITSC 2019.

Guillaume Piolle served as a member of the program committee of the APVP 2019 workshop.

Emmanuelle Anceaume served as a member of the program committee of NCA 2019, EDCC 2019, ICDCS 2019, DEBS 2019, TrustCom 2019, BSCT 2019 and Tokenomics 2019.

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

Michel Hurfin serves as a member of the editorial board of the JISA Journal (Journal of Internet Services and Applications - Springer).

Jean-François Lalande served as a member of the program committee of the IARIA International Journal on Advances in Security.

9.1.3.2. Reviewer - Reviewing Activities

Michel Hurfin served as a reviewer for the International Journal of Control , the Journal on Discrete Event Dynamic Systems - Theory and Applications, and IEEE Transactions on Dependable and Secure Computing.

Jean-François Lalande served as a reviewer for the following international journals: Journal of Cyber Security and Mobility, IEEE Transactions on Reliability, Journal Elsevier FGCS, IEEE Transactions on Industrial Informatics, MDPI Applied Sciences, Journal of Universal Computer Science.

Eric Totel served as a reviewer for the European Congress of Embedded Real Time Software and Systems.

Pierre Wilke served as a reviewer for the 12th International Conference on Security for Information Technology and Communications (SecITC 2019).

Guillaume Hiet served as a reviewer for the Journal of Computer Security and IEEE Design & Test

9.1.4. Invited Talks

Ludovic Mé. *Cyber security: current challenges*. LIRIMA Franco-African webinar. Sept. 2019.

Jean-François Lalande: *Obfuscated Android Application Development*, CECC 2019, Munich.

Emmanuelle Anceaume: *Abstractions for permissionless distributed systems*, SSL seminar, Nancy, 5 December 2019.

Emmanuelle Anceaume: *Can we safely adapt the construction of permissionless blockchains to user demand ?*, *Journées Futur & Rupture*, ParisTech, 31 March 2019.

Emmanuelle Anceaume: *Beyond the block: a lego blockumentary*, SecDays, 9-10 January 2019.

Emmanuelle Anceaume: *Round table at the "Future & Rupture day"* with Mr Ronan Le Gleut (Sénateur), Mr. Gérard Memmi (Télécom ParisTech), Mr. Nicolas KozaKiewicz (Atos Worldline) - moderator Patrick Duvaut, February, 28 2019.

9.1.5. Scientific Expertise

- Ludovic Mé served the HCERES evaluation committee of the LAAS-CNRS research lab.
- Ludovic Mé has served the Scientific Council of the LIRIMA (Laboratoire International de Recherche en Informatique et Mathématiques Appliquées).
- Valérie Viet Triem Tong has participated in the scientific evaluation comity *Global Security and Cybersecurity* (CES 39) of the French Research Agency (ANR).

- Jean-François Lalande is part of the advisory board of the starting european project "SIMARGL" (H2020 project): Secure Intelligent Methods for Advanced Recognition of Malware and Stegomalware.

9.1.6. Research Administration

- Christophe Bidan was a member of a recruitment committee for an assistant professor position at CentraleSupélec, Rennes.
- Ludovic Mé is deputy scientific director of Inria, in charge of the cyber security domain.
- Valérie Viet Triem Tong was a member of a recruitment committee for an assistant professor position at CentraleSupélec, Rennes.
- Valérie Viet Triem Tong was a member of a recruitment committee for an assistant professor position at IMT, Brest.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Team members are involved in initial and continuing education in CentraleSupélec, a french institute of research and higher education in engineering and science and ESIR (Ecole Supérieure d'Ingénieur de Rennes) the graduate engineering school of the University of Rennes 1.

In these institutions,

- Gilles Guette is director of corporate relations at ESIR;
- Jean-François Lalande is responsible of the major program dedicated to information systems security and the special track Infosec of CentraleSupélec ;
- Frédéric Tronel and Valérie Viet Triem Tong share the responsability of the *mastère spécialisé* (post-graduate specialization degree) in Cybersecurity. This education was awarded **best French master degree** in the category “Master Cybersecurity masters and Security of systems” in the Eduniversal master ranking 2019.

The teaching duties are summed up in table 1.

	Licence level	Master level	CS†	Univ. Rennes 1	Initial education	Continuing education	2018 -2019
Emmanuelle Anceaume		✓		✓	✓		10
Christophe Bidan	✓	✓	✓		✓	✓	64
Gilles Guette	✓	✓	✓	✓	✓		400
Guillaume Hiet	✓	✓	✓	✓	✓	✓	328 +34*
Jean-François Lalande	✓	✓	✓		✓	✓	326 +70*
Guillaume Piolle	✓	✓	✓	✓	✓	✓	210
Frédéric Tronel	✓	✓	✓	✓	✓	✓	287
Valérie Viet Triem Tong	✓	✓	✓	✓	✓	✓	229
Pierre Wilke	✓	✓	✓		✓	✓	140

Figure 1.

9.2.2. Supervision

HdR defended in 2019, Emmanuelle Anceaume, *Abstractions for permissionless systems*.

PhD defended in 2019: Aurélien Dupin, *Secure multi-partie computations*, started February 2016, supervised by Christophe Bidan(40%), David Pointcheval (30% - ENS) and Renaud Dubois (30% - Thales).

PhD defended in 2019: Aurélien Palisse, *Analyse et détection de logiciels de rançon*, started in 2015, supervised by Jean-Louis Lanet, Colas Le Guernic (DGA) and Hélène Le Bouder (IMT Atlantique);

PhD defended in 2019: 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 defended in 2019: Pernelle Mensah, *Generation and Dynamic Update of Attack Graphs in Cloud Providers Infrastructures*, started in January 2016, supervised by Eric Totel (25%), Guillaume Piolle (25%), Christine Morin (25% - Myriads Inria project), and Samuel Dubus (25% - Nokia);

PhD defended in 2019: Kevin Bukasa, *Vulnerability analysis of embedded systems against physical attacks* , supervised by Jean-Louis Lanet and Ronan Lashermes (SED Inria);

PhD defended in 2019: Ronny Chevalier, *Detecting and Surviving Intrusions - Exploring New Host-Based Intrusion Detection, Recovery, and Response Approaches*, supervised by Guillaume Hiet (50%), David Plaquin (25% - HP) and Ludovic Mé (25%);

PhD in progress: Alexandre Dey, *Continuous Model Learning for Anomaly Detection In the Presence of Highly Adaptative Cyberattacks*, started in November 2019, supervised by Eric Totel (50%) and Ludovic Mé (50%);

PhD in progress: Leopold Ouairy, *Analyse des vulnérabilités dans des systèmes embarqués*, started in 2017, supervised by Jean-Louis Lanet;

PhD in progress: Mathieu Escouteloup *Micro-architectures Sécurisées*, started in 2018, supervised by Jean-Louis Lanet and Jacques Fournier (CEA);

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: 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: 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: Cedric Herzog, *Simulation d'environnement d'observation afin d'éviter le déploiement de malware sur une station de travail*, started in November 2018, supervised by Jean Louis Lanet (50%), Pierre Wilke (25%) and Valérie Viet Triem Tong (25%);

PhD in progress: Benoit Fournier, *Secure routing in drone swarms*, started in November 2018, supervised by Gilles Guette (50%), Jean Louis Lanet (25%) and Valérie Viet Triem Tong (25%);

PhD in progress: Aimad Berady, *Attacker characterization*, started in November 2018, supervised by Christophe Bidan (25%), Guillaume Carat (25%), Gilles Guette (25%), and Valérie Viet Triem Tong (25%);

PhD in progress: Cyprien Gottstein, *Problématiques de stockage et d'interrogation de très grands graphes répartis dans le contexte de l'internet des objets*, started in October 2018, supervised by Michel Hurfin (50%) and Philippe Raipin Parvedy (50%);

PhD in progress: Tomas Conception Miranda, *Profiling and Visualization Android malware*, started in October 2019, supervised by Jean-François Lalande (34%), Valérie Viet Triem Tong (33%), Pierre Wilke (33%).

9.2.2.1. Supervision of external PhD candidates

PhD in progress: Nicolas Bellec, *Security enhancement in embedded hard real-time systems*, started in October 2019, supervised by Isabelle Puaut (50%), Guillaume Hiet (25%), Frédéric Tronel (25%)

PhD in progress: Kevin Le Bon, *Security enhancement in embedded hard real-time systems*, started in October 2018, supervised by Erven Rohu (30%), Guillaume Hiet (35%), Frédéric Tronel (35%)

9.2.3. Juries

Ludovic Mé was a member of the PhD committee for the following PhD thesis:

Amina Saadaoui, *Formal Techniques for Automatic Detection and Resolution of Security Equipment Misconfigurations*, supervised by Adel Bouhoula and Sihem Guemara, University of Carthage, April 2019.

Amir Wonjiga, *User-Centric Security Monitoring in Cloud Environments*, supervised by Christine Morin and Louis Rilling, University of Rennes 1, May 2019.

Ludovic Mé was a member of the committee for the following HDR defense:

Stephane Mocanu, *Cyberdéfense des infrastructures critiques*, University Grenoble Alpes community, January 2019.

Nizar Kheir, *From Cyber-secure to Cyber-resilient Computer Systems: The way forward*, University of Paris-Saclay, May 2019.

Valérie Viet Triem Tong was a member of the PhD committee for the following PhD thesis:

Xin Ye Model, *Checking Self Modifying Code*, supervised by Tayssir Touili et Jifeng He, University of Paris, Septembre 2019.

Alexandre Dang, *Compilation Sécurisée pour la Protection de la Mémoire*, supervised by Frédéric Besson and Thomas Jensen, University of Rennes 1, December 2019.

Jean-François Lalande was a member of the PhD committee for the following PhD thesis:

M. Guillaume Averlant, *Contrôle d'accès dynamique et architecture de sécurité pour la protection des applications sous Android*, Université Fédérale Toulouse Midi- Pyrénées, october 2th 2019.

9.3. Popularization

9.3.1. Articles and contents

- In books/journals for the general public

  Dis, c'est quoi là haut dans le ciel ? C'est un Linux, mon petit ![39], Viet Triem Tong Valérie, Fournier Benoît, Fournier Guillaume, Ouairy Leopold, Cotret Pascal, Guette Gilles, MISC numéro 104, juillet 2019.

- Online publications

  La cybersécurité aux multiples facettes. Steve Kremer, Ludovic Mé, Didier Rémy et Vincent Roca. Blog binaire (https://www.lemonde.fr/blog/binaire/), 4 juillet 2019.

- White Papers - white Papers for Popularization

  Ludovic Mé co-authored (with S. Kremer, D. Rémy and V. Roca) Inria's White Book on Cybersecurity [37].

9.3.2. Interventions

- Valérie Viet Triem Tong has participated to the *Vive la Recherche* event to promote research activities to engineering students at CentraleSupelec
- Valérie Viet Triem Tong has been involved in an artistic performances involving actors and researcher at the Digital Tech Conference, December Rennes, France.

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] S. K. BUKASA.*Vulnerability analysis of embedded systems against physical attacks*, Université Rennes 1, July 2019, https://tel.archives-ouvertes.fr/tel-02418822

[2] R. CHEVALIER.*Detecting and Surviving Intrusions: Exploring New Host-Based Intrusion Detection, Recovery, and Response Approaches*, CentraleSupélec, December 2019, https://hal.inria.fr/tel-02417644

[3] P. MENSAH.*Generation and Dynamic Update of Attack Graphs in Cloud Providers Infrastructures*, CentraleSupélec, June 2019, https://hal.inria.fr/tel-02416305

[4] A. PALISSE.*Analysis and detection of ransomware*, Université de Rennes 1 ; Inria, March 2019, https://hal.archives-ouvertes.fr/tel-02415976

Articles in International Peer-Reviewed Journal

[5] F. BESSON, S. BLAZY, P. WILKE.*CompCertS: A Memory-Aware Verified C Compiler using a Pointer as Integer Semantics*, in "Journal of Automated Reasoning", August 2019, vol. 63, n° 2, p. 369-392 [*DOI :* 10.1007/S10817-018-9496-Y], https://hal.inria.fr/hal-02401182

[6] S. HAMADOUCHE, M. MEZGHICHE, J.-L. LANET.*Hiding a fault enabled virus through code construction*, in "Journal of Computer Virology and Hacking Techniques", 2019, p. 1-22 [*DOI :* 10.1007/S11416-019-00340-Z], https://hal.inria.fr/hal-02416015

[7] K. HEYDEMANN, J.-F. LALANDE, P. BERTHOMÉ.*Formally verified software countermeasures for control-flow integrity of smart card C code*, in "Computers and Security", August 2019, vol. 85, p. 202-224 [*DOI :* 10.1016/J.COSE.2019.05.004], https://hal.sorbonne-universite.fr/hal-02123836

[8] C. KIRCHNER, L. MÉ.*Défis de la recherche scientifique en cyber-sécurité*, in "Annales des Mines - Enjeux Numériques", December 2019, p. 1-15, https://hal.inria.fr/hal-02381411

[9] Y. MOCQUARD, B. SERICOLA, E. ANCEAUME.*Probabilistic Analysis of Rumor Spreading Time*, in "INFORMS Journal on Computing", July 2019, p. 1-20 [*DOI :* 10.1287/IJOC.2018.0845], https://hal.archives-ouvertes.fr/hal-01888300

Invited Conferences

[10] P. GRAUX, J.-F. LALANDE, V. VIET TRIEM TONG.*Obfuscated Android Application Development*, in "CECC 2019 - Central European Cybersecurity Conference", Munich, Germany, ACM Press, November 2019, p. 1-6 [*DOI :* 10.1145/3360664.3361144], https://hal-centralesupelec.archives-ouvertes.fr/hal-02305924

[11] J.-L. LANET.*Ransomware can always be detected but at which cost ?*, in "JATNA'04", Oujda, Morocco, November 2019, https://hal.inria.fr/hal-02416070

International Conferences with Proceedings

[12] E. ANCEAUME, A. D. POZZO, R. LUDINARD, M. POTOP-BUTUCARU, S. TUCCI-PIERGIOVANNI.*Blockchain Abstract Data Type*, in "SPAA 2019 - 31st ACM Symposium on Parallelism in Algorithms and Architectures", Phoenix, United States, ACM, June 2019, p. 1-11 [*DOI :* 10.1145/3323165.3323183], https://hal-cnrs.archives-ouvertes.fr/hal-02380364

[13] A. BERADY, V. VIET TRIEM TONG, G. GUETTE, C. BIDAN, G. CARAT.*Modeling the Operational Phases of APT Campaigns*, in "CSCI 2019 - 6th Annual Conf. on Computational Science & Computational Intelligence", Las Vegas, United States, December 2019, p. 1-6, https://hal.inria.fr/hal-02379869

[14] F. BESSON, S. BLAZY, A. DANG, T. JENSEN, P. WILKE.*Compiling Sandboxes: Formally Verified Software Fault Isolation*, in "ESOP 2019 - 28th European Symposium on Programming", Prague, Czech Republic, LNCS, Springer, April 2019, vol. 11423, p. 499-524 [*DOI :* 10.1007/978-3-030-17184-1_18], https://hal.inria.fr/hal-02316189

[15] E. BOESPFLUG, R. GOURIER, J.-L. LANET.*Predicting the Effect of Hardware Fault Injection*, in "IWBIS 2019 - 4th IEEE International Workshop on Big Data and Information Security", Bali, Indonesia, IEEE, October 2019, p. 103-108 [*DOI :* 10.1109/IWBIS.2019.8935864], https://hal.inria.fr/hal-02416062

[16] R. CHEVALIER, S. CRISTALLI, C. HAUSER, Y. SHOSHITAISHVILI, R. WANG, C. KRUEGEL, G. VIGNA, D. BRUSCHI, A. LANZI.*BootKeeper: Validating Software Integrity Properties on Boot Firmware Images*, in "CODASPY 2019 - Conference on Data and Application Security and Privacy", Dallas, United States, ACM Press, March 2019, p. 1-11, https://arxiv.org/abs/1903.12505 [*DOI :* 10.1145/3292006.3300026], https://hal.inria.fr/hal-02066420

[17] R. CHEVALIER, D. PLAQUIN, C. DALTON, G. HIET.*Survivor: A Fine-Grained Intrusion Response and Recovery Approach for Commodity Operating Systems*, in "ACSAC 2019 - 35th Annual Computer Security Applications Conference", San Juan, Puerto Rico, Proceedings of the 35th Annual Computer Security Applications Conference, December 2019, vol. 2019, https://arxiv.org/abs/1912.06863 [*DOI :* 10.1145/3359789.3359792], https://hal.inria.fr/hal-02289315

[18] G. A. DI LUNA, E. ANCEAUME, L. QUERZONI.*Byzantine Generalized Lattice Agreement*, in "Proceedings of the 34th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2020)", New Orleans, Louisiana, United States, IEEE, May 2020, https://hal-cnrs.archives-ouvertes.fr/hal-02380446

[19] A. DURAND, E. ANCEAUME, R. LUDINARD.*STAKECUBE: Combining Sharding and Proof-of-Stake to build Fork-free Secure Permissionless Distributed Ledgers*, in "International conference on networked systems (NETYS)", Marrakesh, Morocco, June 2019, https://hal.archives-ouvertes.fr/hal-02078072

[20] J.-F. LALANDE, P. GRAUX, T. CONCEPCIÓN MIRANDA.*Orchestrating Android Malware Experiments*, in "MASCOTS 2019 - 27th IEEE International Symposium on the Modeling, Analysis, and Simulation of Computer and Telecommunication Systems", Rennes, France, IEEE Computer societry, October 2019, p. 1-2, https://hal-centralesupelec.archives-ouvertes.fr/hal-02305473

[21] J.-F. LALANDE, V. VIET TRIEM TONG, P. GRAUX, G. HIET, W. MAZURCZYK, H. CHAOUI, P. BERTHOMÉ.*Teaching Android Mobile Security*, in "SIGCSE '19 - 50th ACM Technical Symposium on Computer Science Education", Minneapolis, United States, Proceedings of the 50th ACM Technical Symposium on Computer Science Education, ACM Press, February 2019, p. 232-238 [*DOI :* 10.1145/3287324.3287406], https://hal-centralesupelec.archives-ouvertes.fr/hal-01940652

[22] D. LANOE, M. HURFIN, E. TOTEL, C. MAZIERO.*An Efficient and Scalable Intrusion Detection System on Logs of Distributed Applications*, in "SEC 2019 - 34th IFIP International Conference on ICT Systems Security and Privacy Protection", Lisbonne, Portugal, Springer, June 2019, p. 49-63 [*DOI :* 10.1007/978-3-030-22312-0_4], https://hal.inria.fr/hal-02409487

[23] T. LETAN, Y. RÉGIS-GIANAS.*FreeSpec: Specifying, Verifying and Executing Impure Computations in Coq*, in "9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP '20)", Nouvelle-Orléans, United States, January 2020 [*DOI :* 10.1145/3372885.3373812], https://hal.inria.fr/hal-02422273

[24] Y. MOCQUARD, B. SERICOLA, E. ANCEAUME.*Brief: Explicit and Tight Bounds of the Convergence Time of Average-based Population Protocols*, in "SIROCCO 2019 - 26th International Colloquium Structural Information and Communication Complexity", L'Aquila, Italy, Springer, July 2019, p. 1-4 [*DOI :* 10.1007/978-3-030-24922-9_29], https://hal-cnrs.archives-ouvertes.fr/hal-02380422

[25] L. OUAIRY, H. LE BOUDER, J.-L. LANET.*JavaNeighbors: Improving ChuckyJava's neighborhood discovery algorithm*, in "EUSPN 2019 - 10th International Conference on Emerging Ubiquitous Systems and Pervasive Networks", Coimbre, Portugal, November 2019, p. 1-7, https://hal.inria.fr/hal-01950822

[26] V. VIET TRIEM TONG, C. HERZOG, T. CONCEPCIÓN MIRANDA, P. GRAUX, J.-F. LALANDE, P. WILKE.*Isolating malicious code in Android malware in the wild*, in "MALCON 2019 - 14th International Conference on Malicious and Unwanted Software", Nantucket, United States, MALCON 2019, IEEE Computer society, October 2019, https://hal-centralesupelec.archives-ouvertes.fr/hal-02288116

National Conferences with Proceeding

[27] N. OURDI, A. PALISSE, J.-L. LANET.*Classification of ransomwares using Artificial Neural Networks and Bayesian Networks*, in "Third International Conference on Intelligent Computing in Data Sciences", Marrakech, Morocco, November 2019, https://hal.inria.fr/hal-02416086

Conferences without Proceedings

[28] E. ANCEAUME, Y. BUSNEL, V. CAZACU.*L'art d'extraire des éléments du top-k en temps réel sur des fenêtres glissantes réparties*, in "ALGOTEL 2019 - 21èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Saint Laurent de la Cabrerisse, France, June 2019, p. 1-4, https://hal-imt-atlantique.archives-ouvertes.fr/hal-02118367

[29] E. ANCEAUME, A. GUELLIER, R. LUDINARD, B. SERICOLA.*Sycomore : un registre de transactions distribué et public au débit auto-adaptatif*, in "ALGOTEL 2019 - 21èmes Rencontres Francophones sur les

Aspects Algorithmiques des Télécommunications", Saint Laurent de la Cabrerisse, France, June 2019, p. 1-4, https://hal.archives-ouvertes.fr/hal-02118385

[30] L. CLAUDEPIERRE, P. BESNIER.*Microcontroller Sensitivity to Fault-Injection Induced by Near-Field Electromagnetic Interference*, in "APEMC 2019 - Asia-Pacific International Symposium on Electromagnetic Compatibility", Sapporo, Japan, June 2019, p. 1-4, https://hal.archives-ouvertes.fr/hal-02313980

[31] B. FOURNIER, G. GUETTE, V. VIET TRIEM TONG, J.-L. LANET.*SEER4US, Secured Energy Efficient Routing for UAV Swarms*, in "WiMob 2019 - 15th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications", Barcelona, France, October 2019, p. 1-6, https://hal.archives-ouvertes.fr/hal-02294744

[32] K. LE BON, B. HAWKINS, E. ROHOU, G. HIET, F. TRONEL.*Plateforme de protection de binaires configurable et dynamiquement adaptative*, in "RESSI 2019 - Rendez-Vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information", Erquy, France, May 2019, p. 1-3, https://hal.inria.fr/hal-02385216

[33] Y. LEMMOU, H. LE BOUDER, J.-L. LANET.*Discriminating Unknown Software Using Distance Model*, in "ICACSIS 2019 - 11th International Conference on Advanced Computer Science and Information Systems", Bali, Indonesia, IEEE, October 2019, https://hal.archives-ouvertes.fr/hal-02352861

[34] R. MOUSSAILEB, C. BERRTI, G. DEBOISDEFFRE, N. CUPPENS, J.-L. LANET.*Watch Out! Doxware on The Way...*, in "CRiSIS 2019 - 14th International Conference on Risks and Security of Internet and Systems", Hammamet, Tunisia, October 2019, https://hal.archives-ouvertes.fr/hal-02313650

[35] R. MOUSSAILEB, N. CUPPENS, J.-L. LANET, H. LE BOUDER.*Ransomware Network Traffic Analysis for Pre-Encryption Alert*, in "FPS 2019 - 12th International Symposium on Foundations & Practice of Security", Toulouse, France, November 2019, https://hal.archives-ouvertes.fr/hal-02313656

[36] C. XOSANAVONGSA, E. TOTEL, O. BETTAN.*Discovering Correlations: A Formal Definition of Causal Dependency Among Heterogeneous Events*, in "EuroS&P 2019 - 4th IEEE European Symposium on Security and Privacy", Stockholm, Sweden, IEEE, June 2019, p. 340-355 [*DOI :* 10.1109/EUROSP.2019.00033], https://hal-imt-atlantique.archives-ouvertes.fr/hal-02363431

Scientific Books (or Scientific Book chapters)

[37] S. KREMER, L. MÉ, D. RÉMY, V. ROCA.*Cybersecurity : Current challenges and Inria's research directions*, Inria white book, Inria, January 2019, n^o^ 3, 172, https://hal.inria.fr/hal-01993308

Research Reports

[38] Y. MOCQUARD, F. ROBIN, B. SERICOLA, E. ANCEAUME.*Average-based Population Protocols : Explicit and Tight Bounds of the Convergence Time*, Irisa, July 2019, https://hal.archives-ouvertes.fr/hal-02178618

Scientific Popularization

[39] V. VIET TRIEM TONG, B. FOURNIER, G. FOURNIER, L. OUAIRY, P. COTRET.*Dis, c'est quoi là haut dans le ciel ? C'est un Linux, mon petit*, in "MISC - Le journal de la sécurité informatique", July 2019, https://hal.archives-ouvertes.fr/hal-02380997

Other Publications

[40] E. ANCEAUME, A. DEL POZZO, R. LUDINARD, M. POTOP-BUTUCARU, S. TUCCI-PIERGIOVANNI.*POSTER: Blockchain Abstract Data Type*, ACM, February 2019, p. 1-2, 24th ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming (PPoPP 2019), Poster [*DOI :* 10.1145/3293883.3303705], https://hal.archives-ouvertes.fr/hal-01988364

[41] E. ANCEAUME, M. PAPATRIANTAFILOU, M. POTOP-BUTUCARU, P. TSIGAS.*Distributed Ledger Register: From Safe to Atomic*, July 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02201472

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

Project-Team DIONYSOS

Creation of the Project-Team: 2009 January 01

Keywords:

Computer Science and Digital Science:

A1.1.9. - Fault tolerant systems
A1.2.2. - Supervision
A1.2.4. - QoS, performance evaluation
A1.2.5. - Internet of things
A1.3.3. - Blockchain
A1.3.4. - Peer to peer
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
A6.1.2. - Stochastic Modeling
A6.2.3. - Probabilistic methods
A6.2.4. - Statistical methods
A9.2. - Machine learning
A9.7. - AI algorithmics

Other Research Topics and Application Domains:

B1.2.1. - Understanding and simulation of the brain and the nervous system
B6.2.1. - Wired technologies
B6.2.2. - Radio technology
B6.2.4. - Optic technology
B6.3.2. - Network protocols
B6.3.3. - Network Management
B6.3.5. - Search engines
B6.4. - Internet of things

1. Team, Visitors, External Collaborators

Research Scientists

Gerardo Rubino [Team leader, Inria, Senior Researcher, HDR]
Raymond Marie [Univ de Rennes I, Emeritus, 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 Maillé [IMT Atlantique, Professor, HDR]

César Viho [Univ de Rennes I, Professor, HDR]
Yves Mocquard [Univ de Rennes 1, Assistant Professor, until Aug 2019]

Technical Staff

Quang Pham Tran Anh [Inria, Engineer]
Federico Sismondi [Univ de Rennes I, Engineer, until April 2019]

PhD Students

Corentin Hardy [Inria and Technicolor, PhD Student, until Apr 2019, granted by CIFRE]
Hamza Ben Ammar [Univ de Rennes I, until Mar 2019]
Ximun Castoreo [Inria, PhD Student]
Mohamed Rahali [Univ de Rennes I and B-COM, PhD Student]
Anouar Rkhami [Inria and Nokia Bell Labs, PhD Student]
Illyyne Saffar [Univ de Rennes I and Nokia Bell Labs, PhD Student, granted by CIFRE]
Imane Taibi [Inria, PhD Student]
Ali Hodroj [Univ de Rennes I and Univ Saint Joseph]
Laura Aspirot [UDELAR, Uruguay, PhD Student]
Jorge Graneri [UDELAR, Uruguay, PhD Student]

Post-Doctoral Fellow

Frédérique Robin [Inria, from Oct 2019]

Visiting Scientists

Jorge Graneri [UDELAR, Uruguay, from Sep 2019 until Oct 2019]
Eduardo Mizraji [UDELAR, Uruguay, Sep 2019]
Marvin Nakayama [New Jersey Institute of Technology, USA, Jun 2019]
Franco Robledo [UDELAR, Uruguay, Feb 2019]
Danhua Shangguan [Institute of Applied Physics and Computational Mathematics, Beijing, China, Sep 2019]

Administrative Assistant

Fabienne Cuyollaa [Inria, Administrative Assistant]

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. Highlights of the Year

5.1.1. Awards

Bruno Tuffin was among the five finalists for the Best Contributed Theoretical Paper Award at the Winter Simulation Conference 2019 for his paper “Randomized Quasi-Monte Carlo for Quantile Estimation” co-authored with Z. Kaplan, Y. Li, M. Nakayama (New Jersey Institute of Technology, USA).

5.1.2. Conference

Yassine Hadjadj-Aoul was General co-chair of the 6th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), Paris, France. The conference took place on December 18-20 2019. Its proceedings will appear in the IEEE xPlore Digital Library.

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

7. New Results

7.1. Performance Evaluation

Participants: Gerardo Rubino, Bruno Sericola.

Fluid Queues. Stochastic fluid flow models and, in particular, those driven by Markov chains, have been intensively studied in the last two decades. Not only they have been proven to be efficient tools to mimic Internet traffic flows at a macroscopic level but they are useful tools in many areas of applications such as manufacturing systems or in actuarial sciences, to cite but a few. We propose in [61] a chapter which focus on such a model in the context of performance analysis of a potentially congested system. The latter is modeled by means of a finite-capacity system whose content is described by a Markov driven stable fluid flow. We describe step-by-step a methodology to compute exactly the loss probability of the system. Our approach is based on the computation of hitting probabilities jointly with the peak level reached during a busy period, both in the infinite and finite buffer case. Accordingly we end up with differential Riccati equations that can be solved numerically. Moreover, we are able to characterize the complete distribution of both the duration of congestion and of the total information lost during such a busy period.

Connecting irreducible and absorbing Markov chains. Irreducible Markov chains in continuous time are the basic tool for instance in performance evaluation (typically, a queuing model), where in a large majority of cases, we are interested in the behavior of the modeled system in steady-state. Most metrics used are based on the stationary distribution of the model, under unicity natural conditions. Absorbing Markov chains, also in continuous time, play the equivalent role in dependability evaluation, because realistic models must have a finite lifetime, which corresponds here to the absorption time of the chain. In this case, the object of interest is this lifetime, steady-state gives no useful information about the system, and most of the used metrics are defined based on that object. In [30] with describe different connections between the two worlds together with some consequences of those relations in both areas, that is, both in performance and in dependability.

Transient analysis of Markov queueing models. Analyzing the transient behavior of a queueing system is much harder than studying its steady state, the difference being basically that of moving from a linear system to a linear differential system. However, a huge amount of efforts has been put on the former problem, from all kinds of points of view: trials to find closed-forms of the main state distributions, algorithms for numerical evaluations, approximations of different types, exploration of other transient metrics than the basic state distributions, etc. In [62] we focus on the first two elements, the derivation of closed-forms for the main transient state distributions, and the development of numerical techniques. The chapter is organized as a survey, and the main goal is to position and to underline the role of the uniformization technique, for both finding closed-forms and for developing efficient numerical evaluation procedures. In some cases, we extend the discussion to other related transient metrics that are relevant for applications.

7.2. Distributed Systems

Participants: Hamza Ben Ammar, Yann Busnel, Yassine Hadjadj-Aoul, Yves Mocquard, Frédérique Robin, Bruno Sericola.

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.

Membership management is a classic and fundamental problem in many use cases. In networking for instance, it is useful to check if a given IP address belongs to a black list or not, in order to allow access to a given server. This has also become a key issue in very large-scale distributed systems, or in massive databases. Formally, from a subset belonging to a very large universe, the problem consists in answering the question "Given any element of the universe, does it belong to a given subset?". Since the access of a perfect oracle answering the question is commonly admitted to be very costly, it is necessary to provide efficient and inexpensive techniques in the context where the elements arrive continuously in a data stream (for example, in network metrology, log analysis, continuous queries in massive databases, etc.). In [36], we propose a simple but efficient solution to answer membership queries based on a couple of Bloom filters. In a nutshell, the idea is to contact the oracle only if an item is seen for the first time. We use a classical Bloom filter to remember an item occurrence. For the next occurrences, we answer the membership query using a second Bloom filter, which is dynamically populated only when the database is queried. We provide theoretical bounds on the false positive and negative probabilities and we illustrate through extensive simulations the efficiency of our solution, in comparison with standard solutions such as classic Bloom filters.

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. We consider recently an application to continuous queries, which are processed by a stream processing engine (SPE) to generate timely results given the ephemeral input data. Variations of input data streams, in terms of both volume and distribution of values, have a large impact on computational resource requirements. Dynamic and Automatic Balanced Scaling for Storm (DABS-Storm) [21] is an original solution for handling dynamic adaptation of continuous queries processing according to evolution of input stream properties, while controlling the system stability. Both fluctuations in data volume and distribution of values within data streams are handled by DABS-Storm to adjust the resources usage that best meets processing needs. To achieve this goal, the DABS-Storm holistic approach combines a proactive auto-parallelization algorithm with a latency-aware load balancing strategy.

Sampling techniques constitute a classical method for detection in large-scale data streams. We have proposed a new algorithm that detects on the fly the k most frequent items in the sliding window model [52]. This algorithm is distributed among the nodes of the system. It is inspired by a recent approach, which consists in associating a stochastic value correlated with the item's frequency instead of trying to estimate its number of occurrences. This stochastic value corresponds to the number of consecutive heads in coin flipping until the first tail occurs. The original approach was to retain just the maximum of consecutive heads obtained by an item, since an item that often occurs will have a higher probability of having a high value. While effective for very skewed data distributions, the correlation is not tight enough to robustly distinguish items with comparable frequencies. To address this important issue, we propose to combine the stochastic approach with a deterministic counting of items. Specifically, in place of keeping the maximum number of consecutive heads obtained by an item, we count the number of times the coin flipping process of an item has exceeded a given threshold. This threshold is defined by combining theoretical results in leader election and coupon collector problems. Results on simulated data show how impressive is the detection of the top-k items in a large range of distributions.

Health Big Data Analysis. The aim of the study was to build a proof-of-concept demonstrating that big data technology could improve drug safety monitoring in a hospital and could help pharmacovigilance professionals to make data-driven targeted hypotheses on adverse drug events (ADEs) due to drug-drug interactions (DDI). In [17], we developed a DDI automatic detection system based on treatment data and laboratory tests from the electronic health records stored in the clinical data warehouse of Rennes academic hospital. We also used OrientDb, a graph database to store informations from five drug knowledge databases and Spark to perform analysis of potential interactions between drugs taken by hospitalized patients. Then, we developed a Machine Learning model to identify the patients in whom an ADE might have occurred because of a DDI. The DDI detection system worked efficiently and the computation time was manageable. The system could be routinely employed for monitoring.

Probabilistic analysis of population protocols. 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 finite-state agents. In [23] we 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. This model 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 the 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 to each node to determine the instants at which interactions occur. The rumor spreading time is thus more realistic since it is the real 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 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.

Among the different problems addressed in the model of population protocols, average-based problems have been studied for the last few years. In these problems, agents start independently from each other with an initial integer state, and at each interaction with another agent, keep the average of their states as their new state. In [45] and [63], using a well chosen stochastic coupling, we considerably improve upon existing results by providing explicit and tight bounds of the time required to converge to the solution of these problems. We apply these general results to the proportion problem, which consists for each agent to compute the proportion of agents that initially started in one predetermined state, and to the counting population size problem, which aims at estimating the size of the system. Both protocols are uniform, i.e., each agent's local algorithm for computing the outputs, given the inputs, does not require the knowledge of the number of agents. Numerical simulations illustrate our bounds of the convergence time, and show that these bounds are tight in the sense that among extensive simulations, numerous ones fit very well with our bounds.

Organizing both transactions and blocks in a distributed ledger. We propose in [53] a new way to organize both transactions and blocks in a distributed ledger to address the performance issues of permissionless ledgers. In contrast to most of the existing solutions in which the ledger is a chain of blocks extracted from a tree or a graph of chains, we present a distributed ledger whose structure is a balanced directed acyclic graph of blocks. We call this specific graph a SYC-DAG. We show that a SYC-DAG allows us to keep all the remarkable properties of the Bitcoin blockchain in terms of security, immutability, and transparency, while enjoying higher throughput and self-adaptivity to transactions demand. To the best of our knowledge, such a design has never been proposed.

Performance of caching systems. Several studies have focused on improving the performance of caching systems in the context of Content-Centric Networking (CCN). In [16], we propose a fairly generic model of caching systems that can be adapted very easily to represent different caching strategies, even the most advanced ones. Indeed, the proposed model of a single cache, named MACS, which stands for Markov chain-based Approximation of CCN Caching Systems, can be extended to represent an interconnection of caches under different schemes. In order to demonstrate the accuracy of our model, we proposed to derive models of the two most effective techniques in the literature, namely LCD and LRU-K, which may adapt to changing patterns of access.

One of the most important concerns when dealing with the performance of caching systems is the static or dynamic (on-demand) placement of caching resources. This issue is becoming particularly important with the upcoming advent of 5G. In [33] we propose a new technique exploiting the model previously proposed model in [16], in order to achieve the best trade-off between the centralization of resources and their distribution, through an efficient placement of caching resources. To do so, we model the cache resources allocation problem as a multi-objective optimization problem, which is solved using Greedy Randomized Adaptive Search Procedures (GRASP). The obtained results confirm the quality of the outcomes compared to an exhaustive search method and show how a cache allocation solution depends on the network's parameters and on the performance metrics that we want to optimize.

7.3. Machine learning

Participants: Yassine Hadjadj-Aoul, Corentin Hardy, Quang Pham Tran Anh, Gerardo Rubino, Bruno Sericola, Imane Taibi, César Viho.

Distributed deep learning on edge-devices. A recently celebrated type of deep neural network is the Generative Adversarial Network (GAN). GANs are generators of samples from a distribution that has been learned; they are up to now centrally trained from local data on a single location. We question in [37] the performance of training GANs using a spread dataset over a set of distributed machines, following a gossip approach shown to work on standard neural networks. This performance is compared to the federated learning distributed method, that has the drawback of sending model data to a server. We also propose a gossip variant, where GAN components are gossiped independently. Experiments are conducted with Tensorflow with up to 100 emulated machines, on the canonical MNIST dataset. The position of the paper is to provide a first evidence that gossip performances for GAN training are close to the ones of federated learning, while operating in a fully decentralized setup. Second, to highlight that for GANs, the distribution of data on machines is critical (i.e., i.i.d. or not). Third, to illustrate that the gossip variant, despite proposing data diversity to the learning phase, brings only marginal improvements over the classic gossip approach.

This work is a part of the thesis [14].

Deep reinforcement learning for network slicing. Recent achievements in Deep Reinforcement Learning (DRL) have shown the potential of these approaches to solve combinatorial optimization problems. However, the Deep Deterministic Policy Gradient algorithm (DDPG), which is one of the most effective techniques, is not suitable to deal with large-scale discrete action space, which is the case of the Virtual Network Function-Forwarding Graph (VNF-FG) placement. To deal with this problem, we propose several enhancements to improve DDPG efficiency [25][47]. The conventional DDPG generates only one action per iteration; thus, it slowly explores the action space especially in a large action space. Thus, we propose to enhance the exploration by considering multiple noisy actions. In order to avoid getting stuck at a local minimum, we propose to multiply the number of critic (for Q-value) neural networks [25]. In order to improve further the exploration, we propose in [47] an evolutionary algorithm to evolve these neural networks in order to discover better ones.

The techniques presented above are generic and can be applied to a variety of problems. To make them even more effective for network slicing problems, we have also proposed to combine them with a proposed First-Fit heuristic that allows for even more interesting results.

Machine learning for Indoor Outdoor detection. Detecting whether a mobile user is indoor or outdoor is an important issue which significantly impacts user behavior contextualization and mobile network resource management. In [59] we investigate hybrid/semi-supervised Deep Learning-based methods for detecting the environment of an active mobile phone user. They are based on both labeled and unlabeled large real radio data obtained from inside the network and from 3GPP signal measurements. We have empirically evaluated the effectiveness of the semi-supervised learning methods using new real-time radio data, with partial ground truth information, gathered massively from multiple typical and diversified locations (indoor and outdoor) of mobile users. We also presented an analysis of such schemes as compared to the existing supervised classification methods including SVM and Deep Learning [57].

Cognition of user behavior can be seen as an efficient tool for automation of future mobile networks. The work presented in [51] deals with the user behaviour modeling. The model includes the prediction of two main features related to mobile user context: the environment and the mobility. We investigate Deep Learning based methods for simultaneously detecting the environment and the mobility state. We empirically evaluate the effectiveness of the proposed techniques using real-time radio data, which has been massively gathered from multiple diversified situations of mobile users.

Predicting the future Perceived Quality level with PSQA. PSQA is a technology developed by Dionysos during a period of several years, whose aim is quantifying the Quality of Experience (more precisely, the Perceived Quality) of an application or service built on the Internet around the transport of audio or video-audio signals. The main properties of PSQA are the its accuracy (indistinguishable from a subjective testing session), the fact that it is fully automatic, with no reference, and able to operate in real time. PSQA is based on supervised learning (the tool learns from subjective testing panels); once trained and validated, it works with no human intervention. In the PSQA project we selected the Random Neural Network tool for the supervised learning associated tasks, after a comparison with the available techniques at the beginning of the project. In [31] we recall all these elements, including the numerical aspects on the optimization side of the learning process, and then, we focus in the current developments where the goal is to predict the Perceived Quality in the close future. This includes the description of the Reservoir Computing models for time series forecasting, and of a tool we proposed, called Echo State Queueing Network, which is a mix between Reservoir Computing and Random Neural Networks.

7.4. Future networks and architectures

Participants: Yassine Hadjadj-Aoul, Gerardo Rubino, Quang Pham Tran Anh, Anouar Rkhami.

Machine learning for network slicing. Network Function Virtualization (NFV) provides a simple and effective mean to deploy and manage network and telecommunications' services. A typical service can be expressed in the form of a Virtual Network Function-Forwarding Graph (VNF-FG). Allocating a VNF-FG is equivalent to placing VNFs and virtual links onto a given substrate network considering resources and quality of service (QoS) constraints. The deployment of VNF-FGs in large-scale networks, such that QoS measures and deployment cost are optimized, is an emerging challenge. Single-objective VNF-FGs allocation has been addressed in existing literature; however, there is still a lack of studies considering multi-objective VNF-FGs allocation. In addition, it is not trivial to obtain optimal VNF-FGs allocation due to its high computational complexity even in the single-objective case. Genetic algorithms (GAs) have proved their ability in coping with multi-objective optimization problems, thus we propose, in [26], a GA-based scheme to solve multi-objective VNF-FGs allocation problem. The numerical results confirm that the proposed scheme can provide near Pareto-optimal solutions within a short execution time.

In [25], we explore the potential of deep reinforcement learning techniques for the placement of VNF-FGs. However, it turns out that even the most well-known learning technique is ineffective in the context of a large-scale action space. In this respect, we propose approaches to find out feasible solutions while improving significantly the exploration of the action space. The simulation results clearly show the effectiveness of the proposed learning approach for this category of problems. Moreover, thanks to the deep learning process, the performance of the proposed approach is improved over time.

The placement of services, as described above, is extremely complex. The issue is even more complex when it comes to placing a service on several non-cooperative domains, where the network operators hide their infrastructure to other competing domains. In [56], we address these problems by proposing a deep reinforcement learning based VNF-FG embedding approach. The results provide insights into the behaviors of non-cooperative domains. They also show the efficiency of the proposed VNF-FG deployment approach having automatic inter-domain load balancing.

Consistent QoS routing in SDN networks. The Software Defined Networking (SDN) paradigm proposes to decouple the control plane (decision-making process) and the data plane (packet forwarding) to overcome the limitations of traditional network infrastructures, which are known to be difficult to manage, especially at scale. Although there are previous works focusing on the problem of Quality of Service (QoS) routing in SDN networks, only few solutions have taken into consideration the network consistency, which reflects the adequacy between the decisions made and the decisions that should be taken. Therefore, we propose, in [19], a network architecture that guarantees the consistency of the decisions to be taken in an SDN network. A consistent QoS routing strategy is, then, introduced in a way to avoid any quality degradation of prioritized traffic, while optimizing resources usage. Thus, we proposed a traffic dispersion heuristic in order to achieve this goal. We compared our approach to several existing framework in terms of best-effort flows average throughput, average video bitrate and video Quality of Experience (QoE). The emulations results, which are performed using the Mininet environment, clearly demonstrate the effectiveness of the proposed methodology that outperforms existing frameworks.

Optical networks. In [20] we attack the so called *Capacity Crunch* crisis announced for optical networks infrastructures. This problem refers to the facts that (i) the transmission capacity of an optical fiber is not limitless, (ii) the bandwidth demand continues to increase exponentially and (iii) the limits are getting dangerously close. The cheapest and shortest-term solution is to increase efficiency, because there are several possibilities to do so. This work is a contribution in that direction. We focus on strongly improving the wavelength assignment procedure by moving to an heterogeneous and flexible process, adapting the dimensioning to the individual users' needs in QoS. In the paper we demonstrate that a non-uniform dimensioning strategy and a tighten QoS provision allows to save significant networks capacity, while simultaneously provisioning to each user the QoS established in its Service Level Agreement.

Survivability of internet services is a significant and crucial challenge in designing future optical networks. A robust infrastructure and transmission protocols are needed to handle such a situation so that the users can maintain communication despite the existence of one or more failed components in the network. For this reason, we present in [40] a generalized approach able to tolerate any failure scenario, to the extent the user can still communicate with the remaining components, where a scenario corresponds to an arbitrary set of links in a non-operational state. To assess the survivability problem, we propose a joint solution to the problems listed next. We show how to find a set of primary routes, a set of alternate routes associated with each failure scenario, and the capacity required on the network to allow communication between all users, in spite of the links' failures, while satisfying for each user a specific predefined quality of service threshold, defined in the Service Level Agreement (SLA). Numerical results show that the proposed approach not only enjoys the advantages of low complexity and ease of implementation but is also able to achieve significant resource savings compared to existing methods. The savings are higher than 30% on single link failures and more than a 100% on two simultaneous link failures scenarios as well as in more complex situations.

Network tomography. Internet tomography studies the inference of the internal network performances from end-to-end measurements. For this problem, Unicast probing can be advantageous due to the wide support of unicast and the easy deployment of unicast probing paths. In [48] we propose two statistical generic methods for the inference of additive metrics using unicast probing. Our solutions give more flexibility in the choice of the collection points placement. Moreover, the probed paths are not limited to specific topologies. Firstly, we propose the k-paths method that extends the applicability of a previously proposed solution called Flexicast for tree topologies. It is based on the Expectation-Maximization (EM) algorithm which is characterized by high computational and memory complexities. Secondly, we propose the Evolutionary Sampling Algorithm (ESA) that enhances the accuracy and the computing time but following a different approach. In [49] we present a different approach, targeted at link metrics inference in an SDN/NFV environment (even if it can be exported outside this field) that we called TOM (Tomography for Overlay networks Monitoring). In such an environment, we are particularly interested in supervising network slicing, a recent tool enabling to create multiple virtual networks for different applications and QoS constraints on a Telco infrastructure. The goal is to infer the underlay resources states from the measurements performed in the overlay structure. We model the inference task as a regression problem that we solve following a Neural Network approach. Since getting labeled data for the training phase can be costly, our procedure generates artificial data instead. By creating a large set of random training examples, the Neural Network learns the relations between the measures done at path and link levels. This approach takes advantage of efficient Machine Learning solutions to solve a classic inference problem. Simulations with a public dataset show very promising results compared to statistical-based methods.We explored mainly additive metrics such as delays or logs of loss rates, but the approach can also be used for non-additive ones such as bandwidth.

7.5. Wireless Networks

Participants: Yann Busnel, Yassine Hadjadj-Aoul, Ali Hodroj, Bruno Sericola, César Viho.

Self-organized UAV-based Supervision and Connectivity. The use of drones has become more widespread in recent years. Many use cases have developed involving these autonomous vehicles, ranging from simple delivery of packages to complex emergency situations following catastrophic events. The miniaturization and very low cost of these machines make it possible today to create large meshes to ensure network coverage in disaster areas, for instance. However, the problems of scaling up and self-organization are still open in these use cases. In [35], we propose a position paper that first presents different new requirements for the deployment of unmanned aerial vehicles (UAV) networks, involving the use of many drones. Then, it introduces solutions from distributed algorithms and real-time data processing to ensure quasi-optimal solutions to the raised problems.

More specifically, providing network services access anytime and anywhere is nowadays a critical issue, especially in disaster emergency situations. A natural response to such a need is the use of autonomous flying drones to help finding survivors and provide network connectivity to the rescue teams. In [34], we propose VESPA, a distributed algorithm using only one-hop information of the drones, to discover targets with unknown location and auto-organize themselves to ensure connectivity between them and the sink in a multi-hop aerial wireless network. We prove that connectivity, termination and coverage are preserved during all stages of our algorithm, and we evaluate the algorithm performances through simulations. Comparison with a prior work shows the efficiency of VESPA both in terms of discovered targets and number of used drones.

Enhancing dynamic adaptive streaming over HTTP for multi-homed users. Mobile video traffic accounted for more than half of all mobile data traffic over the past two years. Due to the limited bandwidth, users demand for high-quality video streaming becomes a challenge, which could be addressed by exploiting the emerging diversity of access network and adaptive video streaming. In [39], a network selection algorithm is proposed for Dynamic Adaptive Streaming over HTTP (DASH),the famous international standard on video streaming, to enhance the received video quality to a “multi-homed user” equipped with multiple interfaces. A Multi-Armed Bandit (MAB) heuristic is proposed for a dynamic selection of the best interface at each step. While the Adaptive Bit rate Rules (ABR) used in DASH allow the video player client to dynamically pick the bit rate level according to the perceived network conditions, at each switching step a quality degradation may occur due to the difference in network conditions of the available interfaces. This paper aims to close this gap by (i) designing a MAB algorithm over DASH for a multi-homed user, (ii) evaluating the proposed mechanism through a test-bed implementation, (iii) extending the classic MAB model and (iv) discussing some open issues.

Vehicular networks. According to recent forecasts, constant population growth and urbanization will bring an additional load of 2.9 billion vehicles to road networks by 2050. This will certainly lead to increased air pollution concerns, highly congested roads putting more strain on an already deteriorated infrastructure, and may increase the risk of accidents on the roads as well. Therefore, to face these issues we need not only to promote the usage of smarter and greener means of transportation but also to design advanced solutions that leverage the capabilities of these means along with modern cities’ road infrastructure to maximize its utility. In [38], we explore novel ways of utilizing inter-vehicle and vehicle to infrastructure communication technology to achieve a safe and efficient lane change manoeuvre for Connected and Autonomous Vehicles (CAVs). The need for such new protocols is due to the risk that every lane change manoeuvre brings to drivers and passengers lives in addition to its negative impact on congestion level and resulting air pollution, if not performed at the right time and using the appropriate speed. To avoid this risk, we design two new protocols; one is built upon and extends an existing one, and aims at ensuring a safe and efficient lane change manoeuvre, while the second is an original solution inspired from the mutual exclusion concept used in operating systems. This latter complements the former by exclusively granting lane change permission in a way that avoids any risk of collision.

7.6. Network Economics

Participants: Bruno Tuffin, Patrick Maillé.

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.

In 2019, we have had a particular focus on network neutrality issues, but trying to look at them from original perspectives, and investigating so-called grey zones not yet addressed in the debate.

What implications of a global Internet with neutral and non-neutral portions? Network neutrality is being discussed worldwide, with different countries applying different policies, some imposing it, others acting against regulation or even repealing it as recently in the USA. The goal of [43] is to model and analyze the interactions of users, content providers, and Internet service providers (ISPs) located in countries with different rules. To do so, we build a simple two-regions game-theoretic model and focus on two scenarios of net neutrality relaxation in one region while it remains enforced in the other one. In a first scenario, from an initial situation where both regions offer the same basic quality, one region allows ISPs to offer fast lanes for a premium while still guaranteeing the basic service; in a second scenario the ISPs in both regions play a game on quality, with only one possible quality in the neutral region, and two in the non-neutral one but with a regulated quality ratio between those. Our numerical experiments lead to very different outcomes, with the first scenario benefiting to all actors (especially the ones in the relaxed-neutrality region) and the second one mainly benefiting mostly to ISPs while Content Providers are worse off, suggesting that regulation should be carefully designed.

Investigating a grey zone: sponsored data. Sponsored data, where content providers have the possibility to pay wireless providers for the data consumed by customers and therefore to exclude it from the data cap, is getting widespread in many countries, but is forbidden in others for concerns of infringing the network neutrality principles. We present in [44] a game-theoretic model analyzing the consequences of sponsored data in presence of competing wireless providers, where sponsoring decided by the content provider can be different at each provider. We also discuss the impact on the proportion of advertising on the displayed content. We show that, surprisingly, the possibility of sponsored data may actually reduce the benefits of content providers and on the other hand increase the revenue of ISPs in competition, with a very limited impact on user welfare.

Search engines, bias, consensus, and search neutrality debate. Different search engines provide different outputs for the same keyword. This may be due to different definitions of relevance, and/or to different knowledge/anticipation of users' preferences, but rankings are also suspected to be biased towards own content, which may prejudicial to other content providers. In [41], we make some initial steps toward a rigorous comparison and analysis of search engines, by proposing a definition for a consensual relevance of a page with respect to a keyword, from a set of search engines. More specifically, we look at the results of several search engines for a sample of keywords, and define for each keyword the visibility of a page based on its ranking over all search engines. This allows to define a score of the search engine for a keyword, and then its average score over all keywords. Based on the pages visibility, we can also define the consensus search engine as the one showing the most visible results for each keyword. We have implemented this model and present an analysis of the results in [41].

7.7. Monte Carlo

Participants: Bruno Tuffin, Gerardo Rubino.

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.

Rare event simulation of regenerative systems. Rare events occur by definition with a very small probability but are important to analyze because of potential catastrophic consequences. In [32], we focus on rare event for so-called regenerative processes, that are basically processes such that portions of them are statistically independent of each other. For many complex and/or large models, simulation is the only tool at hand but it requires specific implementations to get an accurate answer in a reasonable time. There are two main families of rare-event simulation techniques: Importance Sampling (IS) and Splitting. In a first part, we briefly remind them and compare their respective advantages but later (somewhat arbitrarily) devote most of the work to IS. We then focus on the estimation of the mean hitting time of a rarely visited set. A natural and direct estimator consists in averaging independent and identically distributed copies of simulated hitting times, but an alternative standard estimator uses the regenerative structure allowing to represent the mean as a ratio of quantities. We see that in the setting of crude simulation, the two estimators are actually asymptotically identical in a rare-event context, but inefficient for different, even if related, reasons: the direct estimator requires a large average computational time of a single run whereas the ratio estimator faces a small probability computation. We then explain that the ratio estimator is advised when using IS. In the third part, we discuss the estimation of the distribution, not just the mean, of the hitting time to a rarely visited set of states. We exploit the property that the distribution of the hitting time divided by its expectation converges weakly to an exponential as the target set probability decreases to zero. The problem then reduces to the extensively studied estimation of the mean described previously. It leads to simple estimators of a quantile and conditional tail expectation of the hitting time. Some variants are presented and the accuracy of the estimators is illustrated on numerical examples.

In [46], we introduce and analyze a new regenerative estimator. A classical simulation estimator of this class is based on a ratio representation of the mean hitting time, using crude simulation to estimate the numerator and importance sampling to handle the denominator, which corresponds to a rare event. But the estimator of the numerator can be inefficient when paths to the set are very long. We thus introduce a new estimator that expresses the numerator as a sum of two terms to be estimated separately. We provide theoretical analysis of a simple example showing that the new estimator can have much better behavior than the classical estimator. Numerical results further illustrate this.

Randomized Quasi-Monte Carlo for Quantile Estimation. Quantile estimation is a key issue in many application domains, but has been proved difficult to efficiently estimate. In [42], we compare two approaches for quantile estimation via randomized quasi-Monte Carlo (RQMC) in an asymptotic setting where the number of randomizations for RQMC grows large but the size of the low-discrepancy point set remains fixed. In the first method, for each randomization, we compute an estimator of the cumulative distribution function (CDF), which is inverted to obtain a quantile estimator, and the overall quantile estimator is the sample average of the quantile estimators across randomizations. The second approach instead computes a single quantile estimator by inverting one CDF estimator across all randomizations. Because quantile estimators are generally biased, the first method leads to an estimator that does not converge to the true quantile as the number of randomizations goes to infinity. In contrast, the second estimator does, and we establish a central limit theorem for it. Numerical results further illustrate these points.

Reliability analysis with dependent components. In the reliability area, the Marshall-Olkin copula model has emerged as the standard tool for capturing dependence between components in failure analysis. In this model, shocks arise at exponential random times, affecting one or several components, thus inducing a natural correlation in the failure process. However, because the number of parameter of the model grows exponentially with the number of components, the tool suffers from the "curse of dimensionality." These models are usually intended to be applied to design a network before its construction; therefore, it is natural to assume that only partial information about failure behavior can be gathered, mostly from similar existing networks. To construct them, we propose in [22] an optimization approach to define the shock's parameters in the copula, in order to match marginal failures probabilities and correlations between these failures. To deal with the exponential number of parameters of the problem, we use a column-generation technique. We also discuss additional criteria that can be incorporated to obtain a suitable model. Our computational experiments show that the resulting tool produces a close estimation of the network reliability, especially when the correlation between component failures is significant.

The Creation Process is an algorithm that transforms a static network model into a dynamic one. It is the basis of different variance reduction methods designed to make efficient reliability estimations on highly reliable networks in which links can only assume two possible values, operational or failed. In [18] the Creation Process is extended to let it operate on network models in which links can assume more than two values. The proposed algorithm, that we called Multi-Level Creation Process, is the basis of a method, also introduced here, to make efficient reliability estimations of highly reliable stochastic flow networks. The method proposed, which consists in an application of Splitting over the Multi-Level Creation Process, is empirically shown to be accurate, efficient, and robust. This work was the first step towards a way to implement an efficient estimation procedure for the problem of flow reliability analysis. Our first solution in that direction was presented in [54], where not only we could develop a procedure providing a significant variance reduction but that allows a direct extension to the final target, the solution to the same estimation problem in the more general case of models where the components are dependent. The idea is an original way of implementing a splitting procedure that leads simultaneously to these two properties.

Rare events in risk analysis. One of the main tasks when dealing with critical systems (systems where specific classes of failures can deal to human losses, or to huge financial losses) is the ability to quantify the associated risks, which is the door that, when opened, leads to paths towards understanding what can happen and why, and towards capturing the relationships existing between the different parts of the system, with respect to those risks. This is also the necessary preliminary work allowing to evaluate the relative importance of different factors, always from the viewpoint of the considered risks, an important component of any disaster management system. Identifying the dominant ones is important to know which parts of the system we must reinforce. The keynote [29] described different tools available for these tasks, and how they can be used depending on the objectives to reach. The focus was on Monte Carlo techniques, the only available ones in general, because the only ones able to evaluate any kind of system, and how they deal with rare events. It also discussed the main related open research problems. The tutorial [64] is closely related to previous talk, but the presentation explores more in general the estimation problem and the main families of techniques available for its solution (Importance Sampling, and the particular case of Zero-Variance methods, Splitting, Recursive Variance Reduction techniques, etc.).

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 (2016-2019) including a PhD thesis supervision (PhD of Corentin Hardy), done with Technicolor. The starting point of this thesis was 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. The thesis has been defended in April 2019.

8.2. Cifre contract on Personalization for Cognitive Autonomic Networks in 5G

Participant: 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.

8.3. Cifre contract on Resiliency as a Service for 5G networks using Machine Learning

Participants: Sofiene Jelassi, Gerardo Rubino.

The is a Cifre contract including a PhD thesis supervision (PhD of Soumaya Kaada), done with Nokia (Paris). It concerns providing on demand and evolving resiliency schemes over 5G network using advanced machine learning algorithms. It relies on a highly flexible network infrastructure supporting both wired and wireless programmable data planes through a highly-efficient distributed network operating system.

8.4. Bilateral Contract with Industry: Nokia Bell Labs

Participants: Yassine Hadjadj-Aoul, Quang Pham Tran Anh, Anouar Rkhami, Gerardo Rubino.

Gerardo Rubino is the coordinator of the research action “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. Regional Initiatives

- Yann Busnel is a member of the ONCOSHARe project (ONCOlogy bigdata SHARing for Research) funded by Brittany and Pays de la Loire regions, with 280.000 k€ for 24 months.
- Bruno Sericola continues to work on the analysis of fluid queues with Fabrice Guillemin from Orange Labs in Lannion, France.

9.2. National Initiatives

ANR

- Yassine Hadjadj-Aoul, Sofiene Jelassi and Gerardo Rubino are participating at 20% of their time to the IRT BCOM granted by the ANR.
- Yann Busnel is a member of the two following projects: INSHARE granted by the ANR (ANR-15-CE19-0024) and BigClin granted by the LabEx CominLabs (ANR-10-LABX-07-01).

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.

Inria Exploratory Action SNIDE We are leading of the Inria Exploratory Action SNIDE (Search Non neutratlIty DEtection) 2019-2023, involving Dionysos and MIMR (Grenoble).

Search engines play a key role to access content and are accused to bias their results to favor their own services among others. This has led to the sensitive search neutrality debate, similar to the network neutrality debate currently discussed on the role of ISPs. Our goal in this project is to develop and apply a methodology aiming at highlighting a bias and quantifying its impact.

An initial version of our meta-engine (which will be further develop by incorporating outlier detection tests) can be found at https://snide.irisa.fr/.

9.3. European Initiatives

- Bruno Sericola continues to work on the analysis of fluid queues with Marie-Ange Remiche from the university of Namur in Belgium.
- Gerardo Rubino has a long collaboration with Sebastián Basterrech at the VSB-Technical University of Ostrava, Czech Republic, on Machine Learning.

9.4. International Initiatives

9.4.1. Inria International Partners

9.4.1.1. Informal International Partners

We keep a long collaboration in research with the CalPoly at Pomona, USA, on the transient analysis of Markovian models.

9.4.2. Participation in Other International Programs

9.4.2.1. Ecos Sud program

Project "Masc"

Title: Mathematical Algorithms for Semantic Cognition

International Partner (Institution - Laboratory - Researcher):

Universidad de la República (Uruguay) - Biophysics - Eduardo Mizraji, Jorge Graneri

Universidad de la República (Uruguay) - Computer science - Pablo Rodríguez-Bocca

Duration: 2018 – 2020

Start year: 2018

MASC is a three-year project (code U17E03) 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, a PhD student whose co-advisors are Prof. Mizraji and G. Rubino from Dionysos, plus Pablo Rodríguez Bocca, from the Engineering Faculty of the university of the Republic. Our contribution to this project is around the use of mathematical tools applied to the analysis of cognition pathologies.

9.4.2.2. Math and Stic AmSud programs

Project "RareDep"

Title: Rare events analysis in multi-component systems with dependent components

International Partner (Institution - Laboratory - Researcher):

Universidad Adolfo Ibañez (Chile) - Faculty of Engineering and Sciences - Javiera Barrera

Universidad de la República Uruguay (Uruguay) - Computer Science - Héctor Cancela

Universidade Federal de Pernambuco (Brazil) - Mathematics - Pablo Martín Rodríguez

Duration: 2019 – 2020

Start year: 2019

See also: http://mansci-web.uai.cl/raredep/RareDep/Welcome.html

The RareDep project focus on developing new techniques addressing two central elements for the improvement of the available tools for risk analysis of complex systems. One is the case of rare events, occurring both in performance and in dependability evaluation of systems modeled as made of many components. Rare events preclude the use of Monte Carlo techniques when the event of interest has a small probability of occurring, and specific methods are necessary, with many open problems in the area. Independence is the usual assumption when building models (more precisely, in almost all works in the field make this assumption), but we know that the assumption is almost never satisfied. We often are constrained by the necessity of assuming independent components in order to be able to use the available methods. In RareDep, we intend to address both problems simultaneously. This needs to develop new variance reduction techniques, for instance in the Importance Sampling family, or in the Splitting one, to be able to exploit data concerning dependencies between the components of the models. This will be built on top of our accumulated experience in the Monte Carlo area (and related fields, such as Quasi-Monte Carlo, numerical integration, etc.), and a starting effort to begin the exploration of what happens when we relax the omnipresent independence hypothesis. We will also explore what happens if we consider new ideas (several coming from the participants of the proposal) for defining new metrics in some specific areas. In these cases, everything is to be done: procedures to deal with rare events, modeling techniques to deal with dependencies between the system's components, and then, both issues at the same time. Our main application area will concern different types of modern networks (in communications, or in energy distribution, for instance).

Project "ACCON"

Title: Algorithms for the Capacity Crunch problem in Optical Networks

International Partner (Institution - Laboratory - Researcher):

Universidad de la República Uruguay (Uruguay) - Computer Science - Héctor Cancela

UTFSM (Chile) - Télématica - Reinaldo Vallejos

Universidad de Valparaiso (Chile) - Computer Science - Marta Barría

Duration: 2019 – 2020

Start year: 2019

See also: http://accon.elo.usm.cl/

The rapid increase in demand for bandwidth from existing networks has caused a growth in the use of telecommunications technologies, especially WDM optical networks. So far, communication technologies have been able to meet the bandwidth demand. Nevertheless, this decade researchers have anticipated a coming "Capacity Crunch" potential problem associated with these networks. It refers to fact that the transmission capacity limit on optical fibers is close to be reached in the near future. It is then urgent to make the current network architectures evolve, in order to satisfy the relentless exponential growth in bandwidth demand. In other words, the performance bottleneck for optical infrastructures is concentrated around this limiting situation, and the most efficient way of preparing the future of these fundamental technological systems that support the backbone of the Internet is to focus on solving the related management problems. In the previously described scientific context, the ACCON project has a main scientific goal: the development of new strategies capable to provide better resource management techniques to face the threat of the Capacity Crunch. To this end, we will explore the utilization of different analytical techniques to evaluate the performance of several network architecture paradigms, in order to assess their viability in the near future. This will provide us the needed insight leading to finding new strategies for efficiently managing the network resources, and consequently, to contribute addressing this coming Capacity Crunch problem.

9.4.2.3. PHC Ulysses

Project "AFFINE"

Title: Achieving Energy Efficient Communication in Future Networks by Supporting Multi-Access Edge Computing in Internet of Things (IoT)

International Partners (Institution - Laboratory - Researcher):

University College Dublin (Ireland) - Computer Science - Lina Xu

Duration: 1 year

Start year: January 2019

Yassine Hadjadj-Aoul and Lina Xu received a grant from the PHC Ulysses (for French-Irish collaboration). The aim of this project is to improve the energy efficiency for data transmission and communication in IoT networks and therefore to reduce electricity consumption and CO_2 emissions.

Yann Busnel has taken part in several events to develop Indo-French collaborations, notably within the framework of Campus France. In particular, he led the round table on Artificial Intelligence and Mathematics at the Knowledge Summit 2 in Lyon in October 2019, in the presence of the Minister, Frédérique Vidal.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

We have received the following international scientists:

- M. Nakayama (New Jersey Institute of Technology, USA): one week in July 2019.
- DanHua ShangGuan (Institute of Applied Physics and Computational Mathematics, Beijing, China), one month in September 2019.
- Vamsi Bulusu from VJTI Mumbai visited us for 4 months between August and Novembre 2019.
- Jorge Graneri (Sep.–Oct.) and Eduardo Mizraji (Sep.), UDELAR, Uruguay, in the context of the MASC project.
- Nicolás Jara, UTFSM, Chile, Dec., in the context of the ACCON project.
- Franco Robledo, UDELAR, Uruguay, in Feb.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

We are organizing the mini-symposium entitled “Probabilistic approaches for studying blockchain dynamics” of the 8th European Congress of Mathematics, 5 - 11 July 2020, Portorož, Slovenia This mini-symposium aims at bringing together international researchers active on blockchain analysis, and allowing the interaction between different expertises: Markov chain, game theory, fluid limits as well as on different types of stochastic processes. It also aims at gather mathematicians using a theoretical approach to analyze blockchain with those characterizing blockchains through simulations or numerical techniques, and thus it will encourage new collaborations.

10.1.1.1. General Chair, Scientific Chair

- Yassine Hadjadj-Aoul was the General Chair for the *The 6th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM’2019)*, Co-sponsored by IEEE, in Paris (France), December 2019.

10.1.1.2. Member of the Organizing Committees

- Gerardo Rubino and Bruno Tuffin are members of the Steering Committee of the International Workshop on Rare Event Simulation (RESIM) and of Monte Carlo Method Conference (MCM) series.
- Yassine Hadjadj-Aoul is member of the Steering Committee of the International Conference on Information and Communication Technologies for Disaster Management (ICT-DM). He is co-chairing the Steering Committee since 2016.

10.1.2. Scientific Events: Selection

10.1.2.1. Chair of Conference Program Committees

- Bruno Tuffin was TPC co-chair of the 14th Workshop on the Economics of Networks, Systems and Computation (NetEcon 2019), June 2019, Phoenix, AZ, USA.
- Yassine Hadjadj-Aoul was TPC co-chair of *The 11th Wireless Days Conference (WD’2019*, Co-sponsored by IEEE, Manchester, April 2019, Manchester, UK.
- Yassine Hadjadj-Aoul was TPC track chair of the “Satellites IoT and M2M Networks” track, for *The 2nd International Conference on Smart Communications and Networking (SmartNets 2019)*, Co-sponsored by IEEE, December 2019, Sharm El Sheik, Egypt.

- Yassine Hadjadj-Aoul was TPC track chair of the “Cloud & Fog Computing” track, for *International Symposium on Networks, Computers and Communications (ISNCC 2019)*, Co-sponsored by IEEE, June 2019, Istanbul, Turkey.

10.1.2.2. Member of the Conference Program Committees

Bruno Tuffin was TPC member of the following conferences:

- ValueTools 2019, Palma de Mallorca, Spain, March 12-15, 2019
- First IEEE Workshop on the ECOnomics of Fog, Edge, and Cloud Computing (ECOFEC 2019), in conjunction with IEEE INFOCOM 2019, Paris, France
- IEEE International Conference on Communications (IEEE ICC 2019), May 20-24 2019, Shanghai, China.
- 12th International Conference on Monte Carlo Methods and Application (MCM 2019), Sydney, Australia, July 8-12, 2019.
- 31th International Teletraffic Congress (ITC 31), 27-29 August 2019, Budapest, Hungary.
- The International Symposium on Ubiquitous Networking (UNet’19), September 16-19, 2019, Limoges, France.
- GECON 2019 - 16th International Conference on Economics of Grids, Clouds, Software, and Services, Leeds, UK, 17-29 Sept. 2019
- The 2019 IEEE 2nd 5G World Forum (5GWF’19), Dresden, Germany, Sept 30-Oct 2, 2019.
- Financial Risk Track, Winter Simulation Conference (WSC’19), National Harbor, MD, USA, December 8-11, 2019.

Patrick Maillé was TPC member of the following conferences:

- ACM MobiHoc 2019, 20th International Symposium on Mobile Ad Hoc Networking and Computing, Catania, Italy, July 2-5, 2019
- IFIP WIOPT 2019, 17th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, Avignon, France, June 3–7, 2019
- IEEE International Conference on Communications (IEEE ICC 2019), May 20-24 2019, Shanghai, China.
- Algotel 2019, Narbonne, France, June 3-7, 2019
- NetEcon 2019, 14th workshop on the Economics of Networks, Systems and Computation, Phoenix, AZ, USA, June 28, 2019
- GIIS 2019, Global Information Infrastructure Symposium, Paris, France, December 18-20, 2019
- AdHoc-Now 2019, 18th International Conference on Ad Hoc Networks and Wireless, Luxembourg, October 1-3, 2019
- ValueTools 2019, Palma de Mallorca, Spain, March 12-15, 2019

Yann Busnel was a member of the program committee of the following events:

- NCA 2019: 18th IEEE International Symposium on Network Computing and Applications, Boston, USA, September 2019.
- AdHoc Now 2019: 18th International Conference on Ad Hoc Networks and Wireless, Luxembourg, October 2019.

Bruno Sericola was TPC member of the following conference:

- ASMTA 2019 : 25th International Conference on Analytical and Stochastic Modelling Techniques and Applications, Moscow, Russia, October 23 — 25, 2019.

Yassine Hadjadj-Aoul was a member of the program committee of the following events (partial list):

- IEEE Globecom 2019 - IEEE Global Communications Conference, Waikoloa, HI, USA, December 9-13, 2019.
- IEEE ICC 2018 - IEEE International Conference on Communications, Shanghai, China, MO, USA, May 20-24, 2019.
- IEEE WCNC 2019 - IEEE Wireless Communications and Networking Conference, Marrakech, Morocco, April 15-19, 2019.
- IEEE CCNC 2019 - IEEE Consumer Communications & Networking Conference, Las Vegas, USA, January 11-14, 2019.
- IEEE MASCOTS - 2019 - IEEE conference on "Modeling, Analysis, and Simulation On Computer and Telecommunication Systems", Rennes, France, October 22-25, 2019.

Gerardo Rubino was or is a member of the Program Committee of the following events:

- QEST: 16th International Conference on Quantitative Evaluation of SysTems, Glasgow, Scotland, Sep. 2019.
- e-Health: 7th International workshop of e-Health Pervasive Wireless Applications and Services, Lyon, France, Oct. 2019.
- MASCOTS: 28th IEEE International Symposium on the Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, Nice, France, Nov. 2020.
- ICN: 19th International Conference on Networks, Lisbon, Portugal, Feb. 2020.

10.1.2.3. Reviewer

Yann Busnel, Bruno Tuffin, Bruno Sericola, Yassine Hadjadh-Aoul, Gerardo Rubino, and César Viho served as reviewers for several major international conferences.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Bruno Tuffin serves as

- Area Editor (simulation): INFORMS Journal on Computing, since Jan. 2015.
- Associate Editor: ACM Transactions on Modeling and Computer Simulation, since July 2009.

Patrick Maillé serves as

- Associate Editor: Electronic Commerce Research and Applications (Elsevier), since April 2019.
- Associate Editor: IEEE Open Journal of Communications Society, since July 2019.

Bruno Sericola serves as 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.

Yassine Hadjadj-Aoul was a guest editor of a special issue on: "Emerging topics in wireless communications for future smart cities" for the MDPI International Journal.

Gerardo Rubino is Associate Editor of the Journal of Dynamics & Games, https://www.aimsciences.org/journal/2164-6066.

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).
- Bruno Sericola, Bruno Tuffin and Gerardo Rubino served as reviewers for several major international journals.
- Yassine Hadjadj-Aoul served as a reviewer for several major international journals such as: IEEE Journal On Selected Area of Communication *JSAC* "Series on Network Softwarization & Enablers", IEEE Transaction on Vehicular Technologies *TVT*, IEEE Communication Letters, IEEE Transaction on Mobile Computing *TMC*, IEEE Access, Elsevier Future Generation Computer Systems *FGCS*, Journal of Network and Computer Applications *JNCA*.

10.1.4. Invited Talks

Bruno Tuffin gave the following invited talks:

- B. Tuffin. Rare-Event Simulation of Regenerative Systems, Peking University, China, October 10, 2019.
- B. Tuffin. Monte Carlo Simulation and Rare Event Estimation, Mathematical and computational modeling of rare events in complex systems, Recife, Brazil, November 18-20, 2019.

Yann Busnel made several invited and keynote talks in 2019:

- *Blockchain Technology : Synchronization, Consensus and Token Economy*. Cycle of seminars as Invited Professor at VJTI Mumbai, India, September 2019.
- *The need for enlightened sampling in the supervision of large-scale networks*. Keynote at the International Conference on Data Sciences and Engineering, IIT Patna, India, September 2019
- *Large-scale data streaming*. Invited seminar at FEIT, University of Technology Sydney, Australia, August 2019
- *Le besoin d'échantillonnage avisé dans la supervision de réseaux à grande échelle*. Invited seminar at Laboratoire I3S de Sophia-Antipolis, France, April 2019
- *Blockchain opportunities in Europe & Diploma certification using the blockchain @ IMT Atlantique*. Invited talk at International workshop on Blockchain for Global Good, VJTI Mumbai, India, March 2019
- *Le besoin d'échantillonnage avisé dans la supervision de réseaux à grande échelle*. Invited seminar at Laboratoire d'Informatique de Grenoble, France, January 2019.

Yassine Hadjadj-Aoul made several invited and keynote talks in 2019:

- *Reliable Network Function Placement using Evolutionary Algorithms*. Invited tutorial during the 8th IEEE International Conference on Smart Communications in Network & Technologies (SaCoNet), Oran, Algeria, Dec. 2019
- *Virtual Network Function Forwarding Graph Embedding in 5G and post 5G Networks*. Invited Keynote during the 8th IEEE International Conference on Smart Communications in Network & Technologies (SaCoNet), Oran, Algeria, Dec. 2019
- *Efficient support of the upcoming Massive number of IoT devices & in future wireless networks*. Invited talk during the national seminar, SEOC Day, on Embedded Systems and Communicating Objects, Paris France, April 2019

Gerardo Rubino made several invited and keynote talks in 2019:

- *Training recurrent Random Neural Networks: first and second-order techniques, reservoir models, numerical aspects*. Invited talk at ISCIS'19, Conference on Modelling Methods in Computer Systems, Networks and Bioinformatics, Polish Academy of Sciences, Paris, Oct. 15–16, 2019.
- *Relations between irreducible and absorbing Markov chains*. Invited talk at the Fall Western Sectional Meeting of the AMS, Riverside, California, November 9–10, 2019.
- *Rare event analysis in technological catastrophes*. Keynote at the 6th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), Paris, December 18–20, 2019.

Gerardo Rubino gave a seminar at the Computer Science Department of the UDELAR, in Nov. 2019, about Network Tomography. He also gave a 3 hours-tutorial on dependability analysis at the University Adolfo Ibáñez, Santiago, Chile, in Jan. 2019.

10.1.5. Leadership within the Scientific Community

- Bruno Tuffin is an elected member of I-sim council of Informs society, 2018-2020.
- 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 scientific committee of GT ARC (Automatique et Réseaux de Communication) scientific committee (since Nov. 2017).
- Gerardo Rubino belongs to the CSV (the technical committee) of the Images and Networks Cluster of Brittany, France, since its foundation.
- Gerardo Rubino is the coordinator of the research action “Analytics and machine learning”, in collaboration with Nokia Bell Labs.

10.1.6. Research Administration

- Yann Busnel is Head of « Network System, Cybersecurity and Digital law » Research Department at IMT Atlantique (2017–*).
- Yann Busnel is Head of « Network, Telecommunication and Services » Research Department (D2) at IRISA (2019–*).
- 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.
- Bruno Sericola is member of the research commission of the academic council of University of Rennes 1.
- César Viho is member of the HdR commission of the scientific council of Université Rennes 1.
- César Viho is the director of the MathSTIC doctoral school of Université Bretagne Loire.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master: Bruno Tuffin, MEPS (probablility, queuing systems, simulation) , 35 hours, M1, University of Rennes 1, France

Master: Bruno Tuffin, GTA (Game Theory and Applications), 15 hours, M2, University of Rennes 1, France

MOOC on Queuing Theory, available on EdX: Patrick Maillé (in charge of one week of class)

IMT Atlantique 1st year: Yann Busnel, Wireless Autonomous Robot Swarm, 9 hours.

IMT Atlantique 3rd year: Yann Busnel, Big Data and Stream Processing, 9 hours.

Licence: Bruno Sericola, Mathematics, 14h, L2, IUT/University of Rennes 1, 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 M1: César Viho, Networks:Rennes 1 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

Master, 2nd year: Yassine Hadjadj-Aoul, Scalable Network Infrastructure (SNI), 10 hours, The Research in Computer Science (SIF) master and EIT Digital Master/University of Rennes 1, 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

Sofiene Jelassi is the manager of the master program “Heterogeneous Networks and Systems”, Istic/University of Rennes 1, France

Master pro 2nd year: Sofiene Jelassi, Supervision of heterogeneous networks, 32 hours, Istic/University of Rennes 1, France

Master pro 2nd year: Sofiene Jelassi, Cloud & SDN virtualization, 32 hours, Istic/University of Rennes 1, France

Master pro 2nd year: Sofiene Jelassi, Multimedia networks, 32 hours, Istic/University of Rennes 1, France

Bachelor L1: Sofiene Jelassi, Programming Algorithms, 12 hours, SPM/University of Rennes 1, France

Master, 2nd year: Gerardo Rubino, Scalable Network Infrastructure (SNI), 10 hours, The Research in Computer Science (SIF) master and EIT Digital Master/University of Rennes 1, France

Supelec Rennes 3rd year: Gerardo Rubino, Dependability Analysis, 15 hours.

Master 2nd year: Gerardo Rubino, Quality of Experience, 2×4 hours (two different groups of students), Esir/University of Rennes 1, France

Esir/University of Rennes 1, 1st year, Graph theory and algorithms, 20 hours.

10.2.2. Supervision

PhD in progress: Ximun Castoreo, Measurements to check network neutrality, University of Rennes 1, Started 10/2018, supervised by Bruno Tuffin.

PhD in progress: Ayman Chouayakh, Auctions for spectrum allocations in 5G, CIFRE Thesis with Orange Labs/University of Rennes 1, Started 03/2017, supervised by Patrick Maillé.

PhD in progress: Hiba Dakdouk, Multi-player multi-armed bandit problems and applications in IoT, CIFRE Thesis with Orange Labs/University of Rennes 1, Started 01/2019, supervised by Patrick Maillé.

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: Jérôme Henry, “Indoor localization using the most recent standards of 802.11”, IMT Atlantique. Advisors: N. Montavont (IMT Atlantique), Romaric Ludinard (IMT Atlantique), Y. Busnel (IMT Atlantique). Defense in 2021.

PhD in progress: Anouar Rkhami, “Data analytics for optimized resources’ management in future 5G networks”; started on Oct. 2018; Advisors: Gerardo Rubino and Yassine Hadjadj-Aoul, and Abdelkader Outtagarts, Inria – Nokia Bell labs.

PhD in progress: Imane Taibi, “Big data analysis for network monitoring and troubleshooting”; started on Nov. 2017; Advisors: Gerardo Rubino, Chadi Barakat and Yassine Hadjadj-Aoul, Inria.

PhD in progress: Ali Hodroj, Enhancing content delivery to multi-homed users in broadband mobile networks, started in November 2015; supervisors: Bruno Sericola, Marc Ibrahim and Yassine Hadjadj-Aoul, University Rennes 1 and St Joseph University of Beyrouth.

PhD in progress: Mohamed Rahali, “Machine learning-based monitoring and management for hybride SDN networks”; started on Oct. 2017; Advisors: Gerardo Rubino, Inria, and Jean-Michel Sanner, B-COM.

Post-doc in progress: Noël Gillet, “Analyse de flux de données large-échelle pour les données massives en santé”, IMT Atlantique. Advisor: Y. Busnel. October 2018-September 2019.

Post-doc in progress: Frédérique Robin, Modelling the dynamic behaviour of blockchains and analyzing their performance, Inria Rennes. Supervisors: Bruno Sericola and Emmanuelle Anceaume from team Cidre.

PhD: Corentin Hardy, Contribution to the development of deep learning in distributed systems, Defense in April 2019. Supervisors: Bruno Sericola and Erwan Le Merrer from Technicolor.

PhD: Jean-Michel Sanner; Cifre Grant, Orange Labs, “SDN technologies for network services performances adaptation of carriers networks”, Defense in July 2019. Supervisors: Yassine Hadjadj-Aoul and Gerardo Rubino.

PhD: Hamza Ben Ammar, “Network and cache resources optimization for efficient media content delivery in CCN”, Defense in March 2019; advisors: Yassine Hadjadj-Aoul, Gerardo Rubino and Soraya Ait Chellouche, University Rennes 1 [12].

PhD: Laura Aspirot, “Fluid Approximations for Stochastic Telecommunication Models”, University of the Republic, Uruguay. Advisors: E. Mordecki (Uruguay) and G. Rubino (France). Defended on November 21, 2019, [11].

PhD: Jorge Graneri, “Semantic cognition - A mathematical approximation.”, University of the Republic, Uruguay. Advisors: E. Mizraji (Uruguay) and G. Rubino (France). Defended on November 22, 2019, [13].

PhD in progress: Illyyne Saffar, “Personalization for Cognitive Autonomic Networks in 5G”; started on Feb. 2017; Advisors: César Viho (Université Rennes 1) and Kamal Deep-Singh (UJM Saint-Etienne) and Marie Line Alberi-Morel (Nokia Bell labs), Université Rennes 1 and Nokia Bell labs. Defense in June 2020.

10.2.3. Juries

Gerardo Rubino was one of the reviewers of the PhD of Fetia Bannour, Software Defined Networking: Extending SDN control to large-scale networks, University of Paris-East, Paris, Nov. 19, 2019.

Bruno Tuffin was rapporteur for the PhD of Wael Labidi, Smart grid-aware radio engineering in 5G mobile networks. Université de Paris-Saclay, Télécom SudParis, March 2019.

Patrick Maillé was a member of the PhD jury of Lilian Besson, Multi-Players Bandit Algorithms for Internet of Things Networks, Supelec Rennes, November 2019

Yann Busnel was rapporteur of the PhD defense committee of Abderrahmen Kammoun, Université Jean Monnet, Saint- Etienne, July 2019.

Yassine Hadjadj-Aoul was a member of the PhD jury of Ghada Moualla, Resilient Virtualized Network Functions for Data Centers and Decentralized Environments, Côte d'Azur University, September 2019

Bruno Sericola was member of the final selecting board for the recruitment of CNRS researchers in 2019.

César Viho was President of the HDR jury of Geraldine Texier, Vers un Internet programmable offrant des garanties de qualité de service, IMT-Atlantique Rennes, December 2019.

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

Yann Busnel is member of Development Council of Computer Sciences Master at Université de Nantes (2017–*).

10.3.2. Interventions

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] 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^{o} 2, p. 816–825

[3] Y. Hayel, D. Ros, B. Tuffin.*Less-than-Best-Effort Services: Pricing and Scheduling*, in "23rd IEEE Infocom Conference", Hong-Kong, China, March 2004

[4] 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^{o} 3, p. 702–713

[5] P. Maillé, B. Tuffin.*Telecommunication Network Economics: From Theory to Applications*, Cambridge University Press, 2014, 288, https://hal.inria.fr/hal-00908598

[6] S. Mohamed, G. Rubino.*A study of real-time packet video quality using random neural networks*, in "IEEE Trans. Circuits Syst. Video Techn.", 2002, vol. 12, n^{o} 12, p. 1071–1083, https://doi.org/10.1109/TCSVT.2002.806808

[7] H. NABLI, B. SERICOLA.*Performability analysis: a new algorithm*, in "IEEE Transactions on Computers", 1996, vol. 45, n^{o} 4, p. 491–494

[8] B. ORESHKIN, N. RÉGNARD, P. L'ECUYER.*Rate-Based Daily Arrival Process Models with Application to Call Centers*, in "Operations Research", 2016, vol. 64, n^{o} 2, https://hal.inria.fr/hal-01399539

[9] 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

[10] B. TUFFIN.*Bounded Normal Approximation in Highly Reliable Markovian Systems*, in "Journal of Applied Probability", 1999, vol. 36, n^{o} 4

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] L. ASPIROT.*Fluid Approximations for Stochastic Telecommunication Models*, Universidad de la República - Montevideo, December 2019, https://hal.inria.fr/tel-02436267

[12] H. BEN AMMAR.*On models for performance evaluation and cache resources placement in multi-cache networks*, Université Rennes 1, March 2019, https://tel.archives-ouvertes.fr/tel-02109693

[13] J. GRANERI.*Semantic cognition - A mathematical approximation*, Universidad de la República (Montevideo). Facultad de Ciencias, November 2019, https://hal.inria.fr/tel-02436270

[14] C. HARDY.*Contribution to the development of deep learning in distributed systems*, Université Rennes 1, April 2019, https://tel.archives-ouvertes.fr/tel-02284916

[15] J.-M. SANNER.*SDN control plane architecture and network services placement in Telco Infrastructures*, Université de Rennes 1 [UR1], July 2019, https://hal.inria.fr/tel-02443167

Articles in International Peer-Reviewed Journal

[16] H. BEN AMMAR, Y. HADJADJ-AOUL, G. RUBINO, S. AÏT-CHELLOUCHE.*On the performance analysis of distributed caching systems using a customizable Markov chain model*, in "Journal of Network and Computer Applications", March 2019, vol. 130, p. 39-51 [*DOI :* 10.1016/J.JNCA.2019.01.011], https://hal.inria.fr/hal-02427996

[17] G. BOUZILLÉ, C. MORIVAL, R. WESTERLYNCK, P. LEMORDANT, E. CHAZARD, P. LECORRE, Y. BUSNEL, M. CUGGIA.*An Automated Detection System of Drug-Drug Interactions from Electronic Patient Records Using Big Data Analytics*, in "Studies in Health Technology and Informatics", August 2019, vol. 264, p. 45-49 [*DOI :* 10.3233/SHTI190180], https://hal-univ-rennes1.archives-ouvertes.fr/hal-02280252

[18] H. CANCELA, L. MURRAY, G. RUBINO.*Efficient Estimation of Stochastic Flow Network Reliability*, in "IEEE Transactions on Reliability", April 2019, p. 954-970 [*DOI :* 10.1109/TR.2019.2897322], https://hal.inria.fr/hal-02434272

[19] D. E. HENNI, A. GHOMARI, Y. HADJADJ-AOUL. *A consistent QoS routing strategy for video streaming services in SDN networks*, in "International Journal of Communication Systems", October 2019, e4177 [*DOI :* 10.1002/DAC.4177], https://hal.inria.fr/hal-02427661

[20] N. JARA, H. PEMPELFORT, G. RUBINO, R. VALLEJOS. *How much the wavelength dimensioning methods and a tightened QoS provision impact on the dynamic WDM optical networks capacity?*, in "Optical Switching and Networking", January 2020, vol. 35, forthcoming, https://hal.inria.fr/hal-02431514

[21] R. KOTTO-KOMBI, N. LUMINEAU, P. LAMARRE, N. RIVETTI, Y. BUSNEL. *DABS-Storm: A Data-Aware Approach for Elastic Stream Processing*, in "Transactions on Large-Scale Data- and Knowledge-Centered Systems", 2019, vol. 40, p. 58–93 [*DOI :* 10.1007/978-3-662-58664-8_3], https://hal-imt-atlantique.archives-ouvertes.fr/hal-01951682

[22] O. MATUS, J. BARRERA, E. MORENO, G. RUBINO. *On the Marshall–Olkin Copula Model for Network Reliability Under Dependent Failures*, in "IEEE Transactions on Reliability", June 2019, vol. 68, n^{o} 2, p. 451 - 461 [*DOI :* 10.1109/TR.2018.2865707], https://hal.inria.fr/hal-01962918

[23] Y. MOCQUARD, B. SERICOLA, E. ANCEAUME. *Probabilistic Analysis of Rumor Spreading Time*, in "INFORMS Journal on Computing", July 2019, p. 1-20 [*DOI :* 10.1287/IJOC.2018.0845], https://hal.archives-ouvertes.fr/hal-01888300

[24] Q. T. A. PHAM, A. BRADAI, K. D. SINGH, G. PICARD, R. RIGGIO. *Single and Multi-domain Adaptive Allocation Algorithms for VNF Forwarding Graph Embedding*, in "IEEE Transactions on Network and Service Management", March 2019, vol. 16, n^{o} 1, p. 98-112 [*DOI :* 10.1109/TNSM.2018.2876623], https://hal.archives-ouvertes.fr/hal-01904442

[25] Q. T. A. PHAM, Y. HADJADJ-AOUL, A. OUTTAGARTS. *A Deep Reinforcement Learning Approach for VNF Forwarding Graph Embedding*, in "IEEE Transactions on Network and Service Management", December 2019, vol. 16, n^{o} 4, p. 1318-1331 [*DOI :* 10.1109/TNSM.2019.2947905], https://hal.inria.fr/hal-02427641

[26] Q. T. A. PHAM, J.-M. SANNER, C. MORIN, Y. HADJADJ-AOUL. *Virtual network function-forwarding graph embedding: A genetic algorithm approach*, in "International Journal of Communication Systems", August 2019, e4098 [*DOI :* 10.1002/DAC.4098], https://hal.inria.fr/hal-02427993

Invited Conferences

[27] Y. HADJADJ-AOUL. *Efficient support of the upcoming massive number of IoT devices*, in "SEOC Day - Embedded Systems and Communicating Objects", Paris, France, April 2019, https://hal.inria.fr/hal-02443173

[28] Y. HADJADJ-AOUL. *Virtual network function forwarding graph embedding in 5g and post 5g networks*, in "The 8th IEEE International Conference on Smart Communications in Network Technologies (SaCoNet)", Oran, Algeria, December 2019, https://hal.inria.fr/hal-02443170

[29] G. RUBINO. *Rare event analysis in technological catastrophes*, in "ICT-DM 2019 - 6th International Conference on Information and Communication Technologies for Disaster Management", Paris, France, IEEEXplore, December 2019, https://hal.inria.fr/hal-02430575

[30] G. RUBINO. *Relations between irreducible and absorbing Markov chains*, in "2019 - Fall Western Sectional Meeting of the AMS", Riverside, United States, November 2019, 1, https://hal.inria.fr/hal-02430373

[31] G. RUBINO.*Training recurrent Random Neural Networks: first and second-order techniques, reservoir models, numerical aspects*, in "Conference on Modelling Methods in Computer Systems, Networks and Bioinformatics", Paris, France, October 2019, p. 1-51, https://hal.inria.fr/hal-02430427

[32] B. TUFFIN.*Rare-Event Simulation of Regenerative Systems: Estimation of the Mean and Distribution of Hitting Times : Plenary talk*, in "MCM 2019 - 12th International Conference on Monte Carlo Methods and Applications", Sydney, Australia, July 2019, p. 1-61, https://hal.inria.fr/hal-02182946

International Conferences with Proceedings

[33] H. BEN AMMAR, Y. HADJADJ-AOUL, S. AÏT-CHELLOUCHE.*Efficiently allocating distributed caching resources in future smart networks*, in "CCNC 2019 - 16th IEEE Consumer Communications & Networking Conference", Las Vegas, United States, IEEE, January 2019, p. 1-4 [*DOI :* 10.1109/CCNC.2019.8651854], https://hal.inria.fr/hal-01933973

[34] Y. BUSNEL, C. CAILLOUET, D. COUDERT.*Self-organized Disaster Management System by Distributed Deployment of Connected UAVs*, in "ICT-DM 2019 - 6th International Conference on Information and Communication Technologies for Disaster Management", Paris, France, December 2019, p. 1-8, https://hal.inria.fr/hal-02349396

[35] Y. BUSNEL, C. CAILLOUET, D. COUDERT.*Self-organized UAV-based Supervision and Connectivity: Challenges and Opportunities*, in "NCA 2019 - 18th IEEE International Symposium on Network Computing and Applications", Cambridge, United States, IEEE, September 2019, p. 1-5 [*DOI :* 10.1109/NCA.2019.8935060], https://hal.inria.fr/hal-02267396

[36] Y. BUSNEL, N. GILLET.*A Cost-effective and Lightweight Membership Assessment for Large-scale Data Stream*, in "NCA 2019 - 18th IEEE International Symposium on Network Computing and Applications", Cambridge, Massachussets, United States, IEEE International Symposium on Network Computing and Applications (NCA), IEEE, September 2019, p. 1-8, https://hal-imt-atlantique.archives-ouvertes.fr/hal-02270539

[37] C. HARDY, E. LE MERRER, B. SERICOLA.*MD-GAN: Multi-Discriminator Generative Adversarial Networks for Distributed Datasets*, in "IPDPS 2019 - 33rd IEEE International Parallel and Distributed Processing Syposium", Rio de Janeiro, Brazil, IEEE, May 2019, p. 1-12, https://hal.inria.fr/hal-01946665

[38] J. HODGKISS, S. DJAHEL, Y. HADJADJ-AOUL.*An Advanced Coordination Protocol for Safer and more Efficient Lane Change for Connected and Autonomous Vehicles*, in "CCNC 2019 - 16th IEEE Consumer Communications & Networking Conference", Las Vegas, United States, IEEE, January 2019, p. 1-4 [*DOI :* 10.1109/CCNC.2019.8651668], https://hal.inria.fr/hal-01933972

[39] A. HODROJ, M. IBRAHIM, Y. HADJADJ-AOUL, B. SERICOLA.*Enhancing dynamic adaptive streaming over HTTP for multi-homed users using a Multi-Armed Bandit algorithm*, in "2019 15th International Wireless Communications and Mobile Computing Conference (IWCMC)", Tangier, Morocco, IEEE, June 2019, p. 692-697 [*DOI :* 10.1109/IWCMC.2019.8766642], https://hal.inria.fr/hal-02428002

[40] N. JARA, H. PEMPELFORT, G. RUBINO, R. VALLEJOS.*Survivability in optical networks: a solution for the wavelength continuity constraint case*, in "LADC 2019 - 9th Latin-American Symposium on Dependable Computing", Natal, Brazil, Proceedings of the 9th Latin-American Symposium on Dependable Computing, IEEE Xplore, November 2019, p. 1-9, https://hal.inria.fr/hal-02434311

[41] A. KAMOUN, P. MAILLÉ, B. TUFFIN.*Evaluating search engines and defining a consensus implementation*, in "VALUETOOLS 2019 - 12th EAI International Conference on Performance Evaluation Methodologies and Tools", Palma de Majorque, Spain, March 2019, p. 1-10, https://arxiv.org/abs/1808.00958 , https://hal.inria.fr/hal-01852650

[42] Z. T. KAPLAN, Y. LI, M. K. NAKAYAMA, B. TUFFIN.*Randomized Quasi-Monte Carlo for Quantile Estimation*, in "WSC 2019 - Winter Simulation Conference", National Harbor, United States, 2019, p. 1-14, https://hal.inria.fr/hal-02155421

[43] P. MAILLÉ, B. TUFFIN.*Neutral and Non-Neutral Countries in a Global Internet: What Does it Imply?*, in "GECON 2019 - 16th International Conference on the Economics of Grids, Clouds, Systems, and Services", Leeds, United Kingdom, SPRINGER (editor), 2019, p. 1-14, https://hal.inria.fr/hal-02128437

[44] P. MAILLÉ, B. TUFFIN.*Wireless Service Providers Pricing Game in Presence of Possible Sponsored Data*, in "15th International Conference on Network and Service Management (CNSM)", Halifax, Canada, 2019, https://hal.inria.fr/hal-02004692

[45] Y. MOCQUARD, B. SERICOLA, E. ANCEAUME.*Brief: Explicit and Tight Bounds of the Convergence Time of Average-based Population Protocols*, in "SIROCCO 2019 - 26th International Colloquium Structural Information and Communication Complexity", L'Aquila, Italy, Springer, July 2019, p. 1-4 [*DOI :* 10.1007/978-3-030-24922-9_29], https://hal-cnrs.archives-ouvertes.fr/hal-02380422

[46] M. K. NAKAYAMA, B. TUFFIN.*Efficient Estimation of the Mean Hitting Time to a Set of a Regenerative System*, in "WSC 2019 - Winter Simulation Conference", National Harbor, United States, December 2019, p. 1-14, https://hal.inria.fr/hal-02155409

[47] Q. T. A. PHAM, Y. HADJADJ-AOUL, A. OUTTAGARTS.*Evolutionary Actor-Multi-Critic Model for VNF-FG Embedding*, in "CCNC 2020 - IEEE Consumer Communications & Networking Conference", Las Vegas, United States, IEEE, January 2020, p. 1-6, https://hal.inria.fr/hal-02428006

[48] M. RAHALI, J.-M. SANNER, G. RUBINO.*Unicast Inference of Additive Metrics in General Network Topologies*, in "MASCOTS 2019 - 27th IEEE International Symposium on the Modeling, Analysis, and Simulation of Computer and Telecommunication Systems", Rennes, France, IEEE, October 2019, p. 107-115 [*DOI :* 10.1109/MASCOTS.2019.00021], https://hal.inria.fr/hal-02430585

[49] M. RAHALI, J.-M. SANNER, G. RUBINO.*TOM: a self-trained Tomography solution for Overlay networks Monitoring*, in "CCNC 2020 - IEEE Consumer Communications & Networking Conference", Las Vegas, United States, IEEE, January 2020, p. 1-6, https://hal.inria.fr/hal-02434320

[50] I. TAIBI, Y. HADJADJ-AOUL, C. BARAKAT.*When Deep Learning meets Web Measurements to infer Network Performance*, in "CCNC 2020 - IEEE Consumer Communications & Networking Conference", Las Vegas, United States, IEEE, January 2020, p. 1-6, https://hal.inria.fr/hal-02358004

Conferences without Proceedings

[51] M.-L. ALBERI-MOREL, I. SAFFAR, K. D. SINGH, C. VIHO.*Multi-task Deep Learning based Environment and Mobility Detection for User Behavior Modeling*, in "WMLC 2019 - International Workshop on Machine Learning for Communications", Avignon, France, June 2019, p. 1-7, https://hal.archives-ouvertes.fr/hal-02113163

[52] E. Anceaume, Y. Busnel, V. Cazacu.*L'art d'extraire des éléments du top-k en temps réel sur des fenêtres glissantes réparties*, in "ALGOTEL 2019 - 21èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Saint Laurent de la Cabrerisse, France, June 2019, p. 1-4, https://hal-imt-atlantique.archives-ouvertes.fr/hal-02118367

[53] E. Anceaume, A. Guellier, R. Ludinard, B. Sericola.*Sycomore : un registre de transactions distribué et public au débit auto-adaptatif*, in "ALGOTEL 2019 - 21èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Saint Laurent de la Cabrerisse, France, June 2019, p. 1-4, https://hal.archives-ouvertes.fr/hal-02118385

[54] H. Cancela, L. Murray, G. Rubino.*Splitting-Based Method for Network Reliability Estimation*, in "MCM 2019 - 12th International Conference on Monte Carlo Methods and Applications", Sidney, Australia, July 2019, p. 1-24, https://hal.inria.fr/hal-02434314

[55] D. Patel, H. Suslade, J. Rane, P. Prabhu, S. Saluja, Y. Busnel.*KYC As A Service (KASE) - A Blockchain Approach*, in "International Conference on Modelling, Simulation & Intelligent Computing (MoSICom 2020)", Dubai, United Arab Emirates, January 2020, https://hal-imt-atlantique.archives-ouvertes.fr/hal-02441672

[56] Q. T. A. Pham, A. Bradai, K. D. Singh, Y. Hadjadj-Aoul.*Multi-domain non-cooperative VNF-FG embedding: A deep reinforcement learning approach*, in "INFOCOM 2019 - IEEE International Conference on Computer Communications", Paris, France, IEEE, April 2019, p. 1-6, https://hal.archives-ouvertes.fr/hal-02088819

[57] I. Saffar, M.-L. Alberi-Morel, K. D. Singh, C. Viho.*Machine Learning with partially labeled Data for Indoor Outdoor Detection*, in "CCNC 2019 - 16th IEEE Consumer Communications & Networking Conference", Las Vegas, United States, IEEE, January 2019, p. 1-7 [*DOI :* 10.1109/CCNC.2019.8651736], https://hal.archives-ouvertes.fr/hal-02011454

[58] I. Saffar, M.-L. Alberi-Morel, K. D. Singh, C. Viho.*Mobile User Environment Detection using Deep Learning based Multi-Output Classification*, in "WMNC 2019 - 12th IFIP Wireless and Mobile Networking Conference", Paris, France, September 2019, p. 16-23 [*DOI :* 10.23919/WMNC.2019.8881474], https://hal.archives-ouvertes.fr/hal-02437854

[59] I. Saffar, M.-L. Alberi-Morel, K. D. Singh, C. Viho.*Semi-supervised Deep Learning-based Methods for Indoor Outdoor Detection*, in "ICC 2019 - IEEE International Conference on Communications", Shangai, China, IEEE, May 2019, p. 1-7 [*DOI :* 10.1109/ICC.2019.8761297], https://hal.archives-ouvertes.fr/hal-02011449

[60] I. Saffar, K. D. Singh, M.-L. Alberi-Morel, C. Viho.*Deep Learning based Speed Profiling for Mobile Users in 5G Cellular Networks*, in "Globecom - IEEE Global Communications Conference", Waikoloa, HI, United States, IEEE, December 2019, https://hal.archives-ouvertes.fr/hal-02437850

Scientific Books (or Scientific Book chapters)

[61] F. Guillemin, M.-A. Remiche, B. Sericola.*Busy period, congestion analysis and loss probability in fluid queues*, in "Advanced Trends in Queueing Theory", V. Anisimov, N. Limnios (editors), Mathematics and Statistics Series, Sciences, Iste & J. Wiley, London, 2019, https://hal.archives-ouvertes.fr/hal-02422782

[62] G. RUBINO.*Transient analysis of Markovian queueing systems: a survey with focus on closed-forms and uniformization*, in "Advanced Trends in Queueing Theory", V. ANISIMOV, N. LIMNIOS (editors), ISTE @ Wiley, 2020, vol. 2, forthcoming, https://hal.inria.fr/hal-02431271

Research Reports

[63] Y. MOCQUARD, F. ROBIN, B. SERICOLA, E. ANCEAUME.*Average-based Population Protocols : Explicit and Tight Bounds of the Convergence Time*, Irisa, July 2019, https://hal.archives-ouvertes.fr/hal-02178618

Other Publications

[64] G. RUBINO.*Dependability assessment models and evaluation techniques*, January 2019, Lecture - Tutorial, https://hal.inria.fr/hal-02434322

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

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
A1.3.1. - Web
A1.3.6. - Fog, Edge
A2.1.3. - 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
A2.5.5. - Software testing
A2.6.2. - Middleware
A2.6.4. - Ressource management
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
B9.5.1. - Computer science
B9.10. - Privacy

1. Team, Visitors, External Collaborators

Research Scientist

Djamel Eddine Khelladi [CNRS, Researcher]

Faculty Members

Olivier Barais [Team leader, Univ de Rennes I, Professor, HDR]
Mathieu Acher [Univ de Rennes I, Associate Professor]
Arnaud Blouin [INSA Rennes, Associate Professor]
Johann Bourcier [Univ de Rennes I, Associate Professor, HDR]
Benoit Combemale [Univ de Toulouse 1 Capitole, Professor, HDR]

Jean-Marc Jezequel [Univ de Rennes I, Professor, HDR]
Noel Plouzeau [Univ de Rennes I, Associate Professor]

External Collaborator

Gurvan Le Guernic [DGA]

Technical Staff

Amine Benelallam [Inria, Engineer, until Feb 2019]
Emmanuel Chebbi [Inria, Engineer, from Aug 2019 until Sep 2019]
Caroline Landry [Inria, Engineer, until Nov 2019]
Didier Vojtisek [Inria, Engineer]

PhD Students

Alif Akbar Pranata [Inria, PhD Student]
June Benvegnu Sallou [Univ de Rennes I, PhD Student]
Emmanuel Chebbi [Inria, PhD Student, from Oct 2019]
Antoine Cheron [FaberNovel, PhD Student, granted by CIFRE]
Fabien Coulon [Obeo, PhD Student, granted by CIFRE]
Jean-Emile Dartois [Institut de recherche technologique B-com, PhD Student]
Alejandro Gomez Boix [Inria, PhD Student, until Oct 2019]
Pierre Jeanjean [Inria, PhD Student]
Romain Lebouc [KEREVAL, PhD Student, granted by CIFRE]
Manuel Leduc [Univ de Rennes I, PhD Student, until Nov 2019]
Dorian Leroy [TU Wien, PhD Student]
Gauthier Lyan [Keolis Rennes, PhD Student, granted by CIFRE]
Hugo Martin [Univ de Rennes I, PhD Student]
Ludovic Mouline [SnT Luxembourg, PhD Student, until Nov 2019]
Youssou Ndiaye [Orange Labs, PhD Student, until Nov 2019, granted by CIFRE]
Oscar Luis Vera Perez [Inria, PhD Student]

Post-Doctoral Fellow

Juliana Alves Pereira [Univ de Rennes I, Post-Doctoral Fellow]

Visiting Scientists

Nelly Bencomo [Aston University, UK, from Sep 2019]
Sophie Ebersold [Université de Toulouse, Mar 2019]
Paul Temple [University de Namur, until Feb 2019]

Administrative Assistants

Fanny Banor [CNRS, Administrative Assistant, until Mar 2019]
Sophie Maupile [CNRS, Administrative Assistant, from Apr 2019]

2. Overall Objectives

2.1. Overall objectives

DIVERSE's research agenda targets core values of software engineering. In this fundamental domain we focus and develop models, methodologies and theories to address major challenges raised by the emergence of several forms of diversity in the design, deployment and evolution of software-intensive systems. Software diversity has emerged as an essential phenomenon in all application domains born by our industrial partners. These application domains range from complex systems brought by systems of systems (addressed in collaboration with Thales, Safran, CEA and DGA) and Instrumentation and Control (addressed with EDF) to pervasive combinations of Internet of Things and Internet of Services (addressed with TellU and Orange) and tactical information systems (addressed in collaboration with civil security). Today these systems seem to be radically all different, but we envision a strong convergence of the scientific principles that underpin their construction and validation, bringing forwards sane and reliable methods for the design of **flexible and open yet dependable systems**. Flexibility and openness are both critical and challenging software layer properties that must deal with the following four dimensions of diversity: **diversity of languages**, used by the stakeholders involved in the construction of these systems; **diversity of features**, required by the different customers; **diversity of runtime environments**, in which software has to run and adapt; **diversity of implementations**, which are necessary for resilience by redundancy.

In this context, the central 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 yet valid ways, to changes in the requirements and environment. Conversely, explicitly handling diversity is a great opportunity to allow software to spontaneously explore alternative design solutions. Concretely, we want to provide software engineers with the following abilities:

- 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 summarized below:

Scientific objective for DIVERSE: To automatically **compose and synthesize software diversity** from design to runtime to **address unpredictable evolution 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 this aspect stands as 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) [120]. 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 [91]. 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 [73]. 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 in the development of a sound *Software Language Engineering* [0], including a unified typing theory that integrate models as first class entities [123].

[0] See http://planet-sl.org

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. [97] indicates that DSMLs can pave the way for wider industrial adoption of MDE.

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 development and study of appropriate mechanisms for continuous design and runtime 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 runtime behavior is extremely promising [105].

3.1.2. Variability modeling

While the basic vision underlying *Software Product Lines* (SPL) can probably be traced back to David Parnas' seminal article [113] 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 [111]. 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* [110]. Bosch provides a different definition [79]: *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 [85]. 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 [93]. 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 [80]. Variability management is thus growingly seen as the cornerstone of SPL development, covering the entire development life cycle, from requirements elicitation [125] to product derivation [130] to product testing [109], [108].

Halmans *et al.* [93] 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.* [115]: 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.* [103] (feature, hardware platform, performances and attributes variability) or Bass *et al.* [71] who discusses 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* [99]. 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 [80]. 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 [75], [87]. Model Driven Engineering (MDE) makes it possible to automate this weaving process [98]. 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 [106]) manipulating these models to produce for instance a detailed design that can ultimately be transformed to code, or to test suites [114], or other software artifacts.

3.1.3. Component-based software development

Component-based software development [124] 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 [86]: description of requirements on services required from the other components; indirect connections between components thanks to ports and connectors constructs [101]; hierarchical definition of components (assemblies of components can define new component types); connectors supporting various communication semantics [83]; quantitative properties on the services [78].

In recent years component-based architectures have evolved from static designs to dynamic, adaptive designs (e.g., SOFA [83], Palladio [76], Frascati [107]). 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 a stop/redeploy/start process). Dynamic adaptation 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 [100]. 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 runtime, without human intervention, by adapting themselves [84], [127]. 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 [77], dynamic component based systems must separate architectural design and platform compatibility. This requires support for heterogeneity, since platform evolution 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 [118]". In practice, component-based (and/or service-based) platforms offer reflection APIs that make it possible to introspect the system (to determine 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 [95] 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 [92]. Multi Objectives Search based techniques [89] 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 [126] mainly relies on model analysis, constraint solving [88] and search-based reasoning [102]. DIVERSE leverages in particular the applications of model-based testing in the context of highly-configurable systems and [128] interactive systems [104] as well as recent advances based on diversity for test cases selection [96].

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 [112], or Xholon [0]. 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 [82] proposes a common abstract syntax, which represents the description of the model structure. These elements can be decorated using different directors that reflect the application of a specific model of computation on the model element. Metropolis [72] provides modeling elements amenable to semantically equivalent mathematical models. Metropolis offers a precise semantics flexible enough to support different models of computation. ModHel'X [94] 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 [70], recovery blocks [116] and code randomization [74], 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 [121], [119]. Such methods have been used for example to understand the impact of new software development paradigms [81]. 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 [69].

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 axes 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.

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 [120], [91]. 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.

[0]http://www.primordion.com/Xholon/

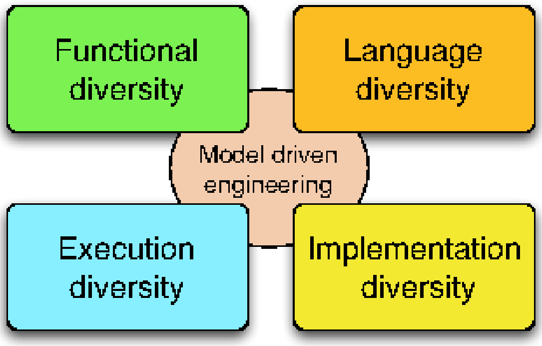


Figure 1. The four research axes of DIVERSE, which rely on a MDE scientific background

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) to set principles for modeling language units that support the modular specification of a modeling language; and (ii) to design mechanisms to assemble these units into 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) gives 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 support to 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. This framework 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 to MoCs and DSMLs design and implementation and to 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 [122] (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 evolution 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 the 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 approaches to deal with the increasing number of concerns in complex systems: 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 reasoning techniques to **automatically analyze** variability models and their interactions with other views on the software intensive system (requirements, architecture, design, code). 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 and learning-based techniques for the exploration of the space of interaction between variability and view models.

We aim at developing 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.

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. Evolution possibilities 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 (e.g., 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 runtime. 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 unforeseen evolution cases.

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 reconfiguration 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 reconfiguration.

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, the number of testing environments to test a software patch or upgrade is the product of the number of execution 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 evolution cases 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 unforeseen 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 evolution, etc. Current fault-tolerance [116] and security [90] 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 rely on implementation 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 and active 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" [117], [129].

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 meta-design principles that maintain diversity in systems and the experimental demonstration of the effects of software diversity. 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 name 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 logic 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. Highlights of the Year

4.1. Highlights of the Year

This year, we would like to highlight the following results:

- In terms of publications among the many articles published this year, articles [37], [28] and [25] have been published at the highest level but above all they represent perfectly the type of research conducted within the team: open research based on studies of major open-source software and in connection with the developer communities.
- We received the "Data Showcase Award" (Figure 2) at the MSR'19 conference (Mining Software Repositories 2019) for the dataset described in the following paper [55] and publicly available on Zenodo (https://zenodo.org/record/1489120).
- Since this year two former PhD students of the team now have a full time researcher position at CNRS: Pierre Laperdrix and Thomas Degueule.
- Four new PhDs and one new HDR have been successfully defended this year.

- A new CNRS junior researcher, Djamel Eddine Khelladi, has joined the team in February 2019. Since his arrival, he submitted a Marie Skłodowska-Curie Action (MSCA) Individual Fellowships (IF), as well as an ANR JCJC in phase 1. Both projects, respectively, CoEvoCCT and MC-Evo[2] are on the topics of software evolution and co-evolution. His research amplifies a new axis around software evolution and maintenance. First results led to the publication at 42nd International Conference on Software Engineering, ICSE, 2020, Seoul, South Korea, an A* top conference in the field of software engineering.

Figure 2. Data Showcase Award, MSR'19

4.1.1. Awards

Most Influential Paper (MIP) award at SLE 2019 https://ins2i.cnrs.fr/fr/cnrsinfo/des-scientifiques-primes-pour-leurs-travaux-sur-les-feature-models

BEST PAPERS AWARDS :

[55]

A. BENELALLAM, N. HARRAND, C. SOTO-VALERO, B. BAUDRY, O. BARAIS.*The Maven Dependency Graph: a Temporal Graph-based Representation of Maven Central*, in "MSR 2019 - 16th International Conference on Mining Software Repositories", Montreal, Canada, ACM, May 2019, p. 344-348 [*DOI :* 10.1109/MSR.2019.00060], https://hal.archives-ouvertes.fr/hal-02080243

5. New Software and Platforms

5.1. 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 detail 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 detail 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.

- Panopticlick https://panopticlick.eff.org/
- BrowserSpy http://browserspy.dk/

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 100,000 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/

5.2. 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

5.3. 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

5.4. Kevoree

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/doc/1.4/frascati-userguide.html).
- 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).

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/

5.5. Melange

KEYWORDS: Model-driven engineering - Meta model - MDE - DSL - Model-driven software engineering - Dedicated langage - Language workbench - Meta-modelisation - Modeling language - 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

5.6. DSpot

KEYWORDS: Software testing - Test amplification

FUNCTIONAL DESCRIPTION: DSpot is a tool that generates missing assertions in JUnit tests. DSpot takes as input a Java project with an existing test suite. As output, DSpot outputs new test cases on console. DSpot supports Java projects built with Maven and Gradle

- Participants: Benoit Baudry, Martin Monperrus and Benjamin Danglot
- Partner: KTH Royal Institute of Technology
- Contact: Benjamin Danglot
- URL: https://github.com/STAMP-project/dspot

5.7. ALE

Action Language for Ecore

KEYWORDS: Meta-modeling - Executable DSML

FUNCTIONAL DESCRIPTION: Main features of ALE include:

- Executable metamodeling: Re-open existing EClasses to insert new methods with their implementations
- Metamodel extension: The very same mechanism can be used to extend existing Ecore metamodels and insert new features (eg. attributes) in a non-intrusive way
- Interpreted: No need to deploy Eclipse plugins, just run the behavior on a model directly in your modeling environment
- Extensible: If ALE doesn't fit your needs, register Java classes as services and invoke them inside your implementations of EOperations.
- Partner: OBEO
- Contact: Benoît Combemale
- URL: http://gemoc.org/ale-lang/

5.8. InspectorGuidget

KEYWORDS: Static analysis - Software testing - User Interfaces

FUNCTIONAL DESCRIPTION: InspectorGuidget is a static code analysing tool. InspectorGuidget analyses UI (user interface/interaction) code of a software system to extract high level information and metrics. InspectorGuidget also finds bad UI coding pratices, such as Blob listener instances. InspectorGuidget analyses Java code.

- Participants: Arnaud Blouin and Benoit Baudry
- Contact: Arnaud Blouin
- Publications: hal-01499106v5 - hal-01308625v2
- URL: https://github.com/diverse-project/InspectorGuidget

5.9. Descartes

KEYWORDS: Software testing - Mutation analysis

FUNCTIONAL DESCRIPTION: Descartes evaluates the capability of your test suite to detect bugs using extreme mutation testing.

Descartes is a mutation engine plugin for PIT which implements extreme mutation operators as proposed in the paper *Will my tests tell me if I break this code?*.

- Participants: Oscar Luis Vera Perez, Benjamin Danglot, Benoit Baudry and Martin Monperrus
- Partner: KTH Royal Institute of Technology
- Contact: Benoit Baudry
- Publications: Descartes: a PITest engine to detect pseudo-tested methods - Tool Demonstration - A Comprehensive Study of Pseudo-tested Methods
- URL: https://github.com/STAMP-project/pitest-descartes

5.10. PitMP

PIT for Multi-module Project

KEYWORDS: Mutation analysis - Mutation testing - Java - JUnit - Maven

FUNCTIONAL DESCRIPTION: PIT and Descartes are mutation testing systems for Java applications, which allows you to verify if your test suites can detect possible bugs, and so to evaluate the quality of your test suites. They evaluate the capability of your test suite to detect bugs using mutation testing (PIT) or extreme mutation testing (Descartes). Mutation testing does it by introducing small changes or faults into the original program. These modified versions are called mutants. A good test suite should able to kill or detect a mutant. Traditional mutation testing works at the instruction level, e.g., replacing ">" by "<=", so the number of generated mutants is huge, as the time required to check the entire test suite. That's why Extreme Mutation strategy appeared. In Extreme Mutation testing, the whole body of a method under test is removed. Descartes is a mutation engine plugin for PIT which implements extreme mutation operators. Both provide reports combining, line coverage, mutation score and list of weaknesses in the source.

- Partners: CSQE - KTH Royal Institute of Technology - ENGINEERING
- Contact: Caroline Landry
- URL: https://github.com/STAMP-project/pitmp-maven-plugin

6. New Results

6.1. Results on Variability modeling and management

In general, we are currently exploring the use of machine learning for variability-intensive systems in the context of VaryVary ANR project https://varyvary.github.io.

6.1.1. Variability and testing.

The performance of software systems (such as speed, memory usage, correct identification rate) tends to be an evermore important concern, often nowadays on par with functional correctness for critical systems. Systematically testing these performance concerns is however extremely difficult, in particular because there exists no theory underpinning the evaluation of a performance test suite, i.e., to tell the software developer whether such a test suite is "good enough" or even whether a test suite is better than another one. This work [37] proposes to apply **Multimorphic testing** and empirically assess the effectiveness of performance test suites of software systems coming from various domains. By analogy with mutation testing, our core idea is to leverage the typical configurability of these systems, and to check whether it makes any difference in the outcome of the tests: i.e., are some tests able to "kill" underperforming system configurations? More precisely, we propose a framework for defining and evaluating the coverage of a test suite with respect to a quantitative property of interest. Such properties can be the execution time, the memory usage or the success rate in tasks performed by a software system. This framework can be used to assess whether a new test case is worth adding to a test suite or to select an optimal test suite with respect to a property of interest. We evaluate several aspects of our proposal through 3 empirical studies carried out in different fields: object tracking in videos, object recognition in images, and code generators.

6.1.2. Variability, sampling, and SAT.

Uniform or near-uniform generation of solutions for large satisfiability formulas is a problem of theoretical and practical interest for the testing community. Recent works proposed two algorithms (namely UniGen and QuickSampler) for reaching a good compromise between execution time and uniformity guarantees, with empirical evidence on SAT benchmarks. In the context of highly-configurable software systems (e.g., Linux), it is unclear whether UniGen and QuickSampler can scale and sample uniform software configurations. We perform a thorough experiment on 128 real-world feature models. We find that UniGen is unable to produce SAT solutions out of such feature models. Furthermore, we show that QuickSampler does not generate uniform samples and that some features are either never part of the sample or too frequently present. Finally, using a case study, we characterize the impacts of these results on the ability to find bugs in a configurable system. Overall, our results suggest that we are not there: more research is needed to explore the cost-effectiveness of uniform sampling when testing large configurable systems. More details [51]. In general, we are investigating sampling algorithms for cost-effectively exploring configuration spaces (see also [63], [67].

6.1.3. Variability and 3D printing.

Configurators rely on logical constraints over parameters to aid users and determine the validity of a configuration. However, for some domains, capturing such configuration knowledge is hard, if not infeasible. This is the case in the 3D printing industry, where parametric 3D object models contain the list of parameters and their value domains, but no explicit constraints. This calls for a complementary approach that learns what configurations are valid based on previous experiences. In this work [41], we report on preliminary experiments showing the capability of state-of-the-art classification algorithms to assist the configuration process. While machine learning holds its promises when it comes to evaluation scores, an in-depth analysis reveals the opportunity to combine the classifiers with constraint solvers.

6.1.4. Variability and video processing.

In an industrial project [24], we addressed the challenge of developing a software-based video generator such that consumers and providers of video processing algorithms can benchmark them on a wide range of video variants. We have designed and developed a variability modeling language, called VM, resulting from the close collaboration with industrial partners during two years. We expose the specific requirements and advanced variability constructs we developed and used to characterize and derive variations of video sequences. The results of our experiments and industrial experience show that our solution is effective to model complex variability information and supports the synthesis of hundreds of realistic video variants. From the software language perspective, we learned that basic variability mechanisms are useful but not enough; attributes and multi-features are of prior importance; meta-information and specific constructs are relevant for scalable and purposeful reasoning over variability models. From the video domain and software perspective, we report on the practical benefits of a variability approach. With more automation and control, practitioners can now envision benchmarking video algorithms over large, diverse, controlled, yet realistic datasets (videos that mimic real recorded videos) – something impossible at the beginning of the project.

6.1.5. Variability and adversarial machine learning

Software product line engineers put a lot of effort to ensure that, through the setting of a large number of possible configuration options, products are acceptable and well-tailored to customers' needs. Unfortunately, options and their mutual interactions create a huge configuration space which is intractable to exhaustively explore. Instead of testing all products, machine learning is increasingly employed to approximate the set of acceptable products out of a small training sample of configurations. Machine learning (ML) techniques can refine a software product line through learned constraints and a priori prevent non-acceptable products to be derived. In this work [53], we use adversarial ML techniques to generate adver- sarial configurations fooling ML classifiers and pinpoint incorrect classifications of products (videos) derived from an industrial video generator. Our attacks yield (up to) a 100% misclassification rate and a drop in accuracy of 5%. We discuss the implications these results have on SPL quality assurance.

6.1.6. Variability, Linux and machine learning

Given a configuration, can humans know in advance the build status, the size, the compilation time, or the boot time of a Linux kernel? Owing to the huge complexity of Linux (there are more than 15000 options with hard constraints and subtle interactions), machines should rather assist contributors and integrators in mastering the configuration space of the kernel. We have developed TuxML https://github.com/TuxML/ an open-source tool based on Docker/Python to massively gather data about thousands of kernel configurations. 200K+ configurations have been automatically built and we show how machine learning can exploit this information to predict properties of unseen Linux configurations, with different use cases (identification of influential/buggy options, finding of small kernels, etc.) The vision is that a continuous understanding of the configuration space is undoubtedly beneficial for the Linux community, yet several technical challenges remain in terms of infrastructure and automation.

Two preprints are available [62] and [49].

A talk has been given at Embedded Linux Conference Europe 2019 (co-located with Open Source Summit 2019) in Lyon about “Learning the Linux Kernel Configuration Space: Results and Challenges” [54].

6.1.7. Variability and machine learning

We gave a tutorial [49] at SPLC 2019 and introduce how machine learning can be used to support activities related to the engineering of configurable systems and software product lines. To the best of our knowledge, this is the first practical tutorial in this trending field. The tutorial is based on a systematic literature review [67] and includes practical tasks (specialization, performance prediction) on real-world systems (VaryLaTeX, x264).

6.2. Results on Software Language Engineering

6.2.1. Software Language Extension Problem

The problem of software language extension and composition drives much of the research in *Software Language Engineering* (SLE). Although various solutions have already been proposed, there is still little understanding of the specific ins and outs of this problem, which hinders the comparison and evaluation of existing solutions. In [34], we introduce the Language Extension Problem as a way to better qualify the scope of the challenges related to language extension and composition. The formulation of the problem is similar to the seminal Expression Problem introduced by Wadler in the late nineties, and lift it from the extensibility of single constructs to the extensibility of groups of constructs, i.e., software languages. We provide a comprehensive definition of the actual constraints when considering language extension, and believe the Language Extension Problem will drive future research in SLE, the same way the original Expression Problem helped to understand the strengths and weaknesses of programming languages and drove much research in programming languages.

6.2.2. A unifying framework for homogeneous model composition

The growing use of models for separating concerns in complex systems has lead to a proliferation of model composition operators. These composition operators have traditionally been defined from scratch following various approaches differing in formality, level of detail, chosen paradigm, and styles. Due to the lack of proper foundations for defining model composition (concepts, abstractions, or frameworks), it is difficult to compare or reuse composition operators. In [33], we stipulate the existence of a unifying framework that reduces all structural composition operators to structural merging, and all composition operators acting on discrete behaviors to event scheduling. We provide convincing evidence of this hypothesis by discussing how structural and behavioral homogeneous model composition operators (i.e., weavers) can be mapped onto this framework. Based on this discussion, we propose a conceptual model of the framework, and identify a set of research challenges, which, if addressed, lead to the realization of this framework to support rigorous and efficient engineering of model composition operators for homogeneous and eventually heterogeneous modeling languages.

6.2.3. 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. Our contribution in [26] is 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.

6.2.4. From DSL specification to interactive computer programming environment

The adoption of Domain-Specific Languages (DSLs) relies on the capacity of language workbenches to automate the development of advanced and customized environments. While DSLs are usually well tailored for the main scenarios, the cost of developing mature tools prevents the ability to develop additional capabilities for alternative scenarios targeting specific tasks (e.g., API testing) or stakeholders (e.g., education). In [47], we propose an approach to automatically generate interactive computer programming environments from existing specifications of textual interpreted DSLs. The approach provides abstractions to complement the DSL specification, and combines static analysis and language transformations to automate the transformation of the language syntax, the execution state and the execution semantics. We evaluate the approach over a representative set of DSLs, and demonstrate the ability to automatically transform a textual syntax to load partial programs limited to a single statement, and to derive a Read-Eval-Print-Loop (REPL) from the specification of a language interpreter.

6.2.5. Live-UMLRT: A Tool for Live Modeling of UML-RT Models

In the context of Model-driven Development (MDD) models can be executed by interpretation or by the translation of models into existing programming languages, often by code generation. In [42] we present Live-UMLRT, a tool that supports live modeling of UML-RT models when they are executed by code generation. Live-UMLRT is entirely independent of any live programming support offered by the target language. This independence is achieved with the help of a model transformation which equips the model with support for, e.g., debugging and state transfer both of which are required for live modeling. A subsequent code generation then produces a self-reflective program that allows changes to the model elements at runtime (through synchronization of design and runtime models). We have evaluated Live-UMLRT on several use cases. The evaluation shows that (1) code generation, transformation, and state transfer can be carried out with reasonable performance, and (2) our approach can apply model changes to the running execution faster than the standard approach that depends on the live programming support of the target language. A demonstration video: https://youtu.be/6GrR-Y9je7Y.

6.2.6. Applying model-driven engineering to high-performance computing: Experience report, lessons learned, and remaining challenges

In [35], we present a framework for generating optimizing compilers for performance-oriented embedded DSLs (EDSLs). This framework provides facilities to automatically generate the boilerplate code required for building DSL compilers on top of the existing extensible optimizing compilers. We evaluate the practicality of our framework by demonstrating a real-world use-case successfully built with it.

6.2.7. Software languages in the wild (Wikipedia)

Wikipedia is a rich source of information across many knowledge domains. Yet, recovering articles relevant to a specific domain is a difficult problem since such articles may be rare and tend to cover multiple topics. Furthermore, Wikipedia's categories provide an ambiguous classification of articles as they relate to all topics and thus are of limited use. In [46], we develop a new methodology to isolate Wikipedia's articles that describe a specific topic within the scope of relevant categories; the methodology uses super- vised machine learning to retrieve a decision tree classifier based on articles' features (URL patterns, summary text, infoboxes, links from list articles). In a case study, we retrieve 3000+ articles that describe software (computer) languages. Available fragments of ground truths serve as an essential part of the training set to detect relevant articles. The results of the classification are thoroughly evaluated through a survey, in which 31 domain experts participated.

6.3. 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 of resources for dynamically adaptive system, one in the field of temporal context model for dynamically adaptive system and a last one to improve the exploration of hidden real-time structures of programming behavior at run time.

6.3.1. Resource-aware models@runtime layer for dynamically adaptive system

In Kevoree, one of the goal is to work on the shipping phases in which we aim at making deployment, and the reconfiguration simple and accessible to a whole development team. This year, we mainly explore two main axes.

In the first one, we try to improve the proposed models that could be used at run time to improve resource usage in two domains: cloud computing [30], [57] and energy [58].

6.3.2. Investigating Machine Learning Algorithms for Modeling SSD I/O Performance for Container-based Virtualization

One of the cornerstones of the cloud provider business is to reduce hardware resources cost by maximizing their utilization. This is done through smartly sharing processor, memory, network and storage, while fully satisfying SLOs negotiated with customers. For the storage part, while SSDs are increasingly deployed in data centers mainly for their performance and energy efficiency, their internal mechanisms may cause a dramatic SLO violation. In effect, we measured that I/O interference may induce a 10x performance drop. We are building a framework based on autonomic computing which aims to achieve intelligent container placement on storage systems by preventing bad I/O interference scenarios. One prerequisite to such a framework is to design SSD performance models that take into account interactions between running processes/containers, the operating system and the SSD. These interactions are complex. In this work [30], we investigate the use of machine learning for building such models in a container based Cloud environment. We have investigated five popular machine learning algorithms along with six different I/O intensive applications and benchmarks. We analyzed the prediction accuracy, the learning curve, the feature importance and the training time of the tested algorithms on four different SSD models. Beyond describing modeling component of our framework, this paper aims to provide insights for cloud providers to implement SLO compliant container placement algorithms on SSDs. Our machine learning-based framework succeeded in modeling I/O interference with a median Normalized Root-Mean-Square Error (NRMSE) of 2.5%.

6.3.3. Cuckoo: Opportunistic MapReduce on Ephemeral and Heterogeneous Cloud Resources

Cloud infrastructures are generally over-provisioned for handling load peaks and node failures. However, the drawback of this approach is that a large portion of data center resources remains unused. In this work [57], we propose a framework that leverages unused resources of data centers, which are ephemeral by nature, to run MapReduce jobs. Our approach allows: i) to run efficiently Hadoop jobs on top of heterogeneous Cloud resources, thanks to our data placement strategy, ii) to predict accurately the volatility of ephemeral resources, thanks to the quantile regression method, and iii) for avoiding the interference between MapReduce jobs and co-resident workloads, thanks to our reactive QoS controller. We have extended Hadoop implementation with our framework and evaluated it with three different data center workloads. The experimental results show that our approach divides Hadoop job execution time by up to 7 when compared to the standard Hadoop implementation. In [44], we presented a demo that leverages unused but volatile Cloud resources to run big data jobs. It is based on a learning algorithm that accurately predicts future availability of resources to automatically scale the ran jobs. We also designed a mechanism that avoids interference between the Big data jobs and co-resident workloads. Our solution is based on Open-Source components such as kubernetes and Apache Spark.

6.3.4. Leveraging cloud unused resources for Big data application while achieving SLA

Companies are more and more inclined to use collaborative cloud resources when their maximum internal capacities are reached in order to minimize their TCO. The downside of using such a collaborative cloud, made of private clouds' unused resources, is that malicious resource providers may sabotage the correct execution of third-party-owned applications due to its uncontrolled nature. In this work [43], we propose an approach that allows sabotage detection in a trustless environment. To do so, we designed a mechanism that (1) builds an application fingerprint considering a large set of resources usage (such as CPU, I/O, memory) in a trusted environment using random forest algorithm, and (2) an online remote fingerprint recognizer that monitors application execution and that makes it possible to detect unexpected application behavior. Our approach has been tested by building the fingerprint of 5 applications on trusted machines. When running these applications on untrusted machines (with either homogeneous, heterogeneous or unspecified hardware from the one that was used to build the model), the fingerprint recognizer was able to ascertain whether the execution of the application is correct or not with a median accuracy of about 98% for heterogeneous hardware and about 40% for the unspecified one.

6.3.5. Benefits of Energy Management Systems on local energy efficiency, an agricultural case study

Energy efficiency is a concern impacting both ecology and economy. Most approaches aiming at reducing the energy impact of a site focus on only one specific aspect of the ecosystem: appliances, local generation or energy storage. A trade-off analysis of the many factors to consider is challenging and must be supported by tools. This work proposes a Model-Driven Engineering approach mixing all these concerns into one comprehensive model [58]. This model can then be used to size either local production means, either energy storage capacity and also help to analyze differences between technologies. It also enables process optimization by modeling activity variability: it takes the weather into account to give regular feedback to the end user. This approach is illustrated by simulation using real consumption and local production data from a representative agricultural site. We show its use by: sizing solar panels, by choosing between battery technologies and specification and by evaluating different demand response scenarios while examining the economic sustainability of these choices.

6.4. 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) observations of software diversity mainly through browser fingerprinting, and (ii) software testing to study and assess the validity of software.

6.4.1. Privacy and Security

6.4.1.1. A Collaborative Strategy for Mitigating Tracking through Browser Fingerprinting

Browser fingerprinting is a technique that collects information about the browser configuration and the environment in which it is running. This information is so diverse that it can partially or totally identify users online. Over time, several countermeasures have emerged to mitigate tracking through browser fingerprinting. However, these measures do not offer full coverage in terms of privacy protection, as some of them may introduce inconsistencies or unusual behaviors, making these users stand out from the rest. In this work [45], we address these limitations by proposing a novel approach that minimizes both the identifiability of users and the required changes to browser configuration. To this end, we exploit clustering algorithms to identify the devices that are prone to share the same or similar fingerprints and to provide them with a new non-unique fingerprint. We then use this fingerprint to automatically assemble and run web browsers through virtualization within a docker container. Thus all the devices in the same cluster will end up running a web browser with an indistinguishable and consistent fingerprint.

6.4.2. Software Testing

6.4.2.1. A Snowballing Literature Study on Test Amplification

The adoption of agile development approaches has put an increased emphasis on developer testing, resulting in software projects with strong test suites. These suites include a large number of test cases, in which developers embed knowledge about meaningful input data and expected properties in the form of oracles. This work [29] surveys various works that aim at exploiting this knowledge in order to enhance these manually written tests with respect to an engineering goal (e.g., improve coverage of changes or increase the accuracy of fault localization). While these works rely on various techniques and address various goals, we believe they form an emerging and coherent field of research, which we call ‘test amplification’. We devised a first set of papers from DBLP, looking for all papers containing ‘test’ and ‘amplification’ in their title. We reviewed the 70 papers in this set and selected the 4 papers that fit our definition of test amplification. We use these 4 papers as the seed for our snowballing study, and systematically followed the citation graph. This study is the first that draws a comprehensive picture of the different engineering goals proposed in the literature for test amplification. In particular, we note that the goal of test amplification goes far beyond maximizing coverage only. We believe that this survey will help researchers and practitioners entering this new field to understand more quickly and more deeply the intuitions, concepts and techniques used for test amplification.

6.4.2.2. Automatic Test Improvement with DSpot: a Study with Ten Mature Open-Source Projects

In the literature, there is a rather clear segregation between manually written tests by developers and automatically generated ones. In this work, we explore a third solution: to automatically improve existing test cases written by developers. We present the concept, design, and implementation of a system called DSpot, that takes developer-written test cases as input (Junit tests in Java) and synthesizes improved versions of them as output. Those test improvements are given back to developers as patches or pull requests, that can be directly integrated in the main branch of the test code base. In this work [28], we have evaluated DSpot in a deep, systematic manner over 40 real-world unit test classes from 10 notable and open-source software projects. We have amplified all test methods from those 40 unit test classes. In 26/40 cases, DSpot is able to automatically improve the test under study, by triggering new behaviors and adding new valuable assertions. Next, for ten projects under consideration, we have proposed a test improvement automatically synthesized by DSpot to the lead developers. In total, 13/19 proposed test improvements were accepted by the developers and merged into the main code base. This shows that DSpot is capable of automatically improving unit-tests in real-world, large-scale Java software.

6.4.2.3. Leveraging metamorphic testing to automatically detect inconsistencies in code generator families

Generative software development has paved the way for the creation of multiple code generators that serve as a basis for automatically generating code to different software and hardware platforms. In this context, the software quality becomes highly correlated to the quality of code generators used during software development. Eventual failures may result in a loss of confidence for the developers, who will unlikely continue to use these generators. It is then crucial to verify the correct behaviour of code generators in order to preserve software quality and reliability. In this work [25], we leverage the metamorphic testing approach to automatically detect inconsistencies in code generators via so-called “metamorphic relations”. We define the metamorphic relation (i.e., test oracle) as a comparison between the variations of performance and resource usage of test suites running on different versions of generated code. We rely on statistical methods to find the threshold value from which an unexpected variation is detected. We evaluate our approach by testing a family of code generators with respect to resource usage and performance metrics for five different target software platforms. The experimental results show that our approach is able to detect, among 95 executed test suites, 11 performance and 15 memory usage inconsistencies.

6.4.3. Software Co-evolution

6.4.3.1. An Empirical Study on the Impact of Inconsistency Feedback during Model and Code Co-changing

Model and code co-changing is about the coordinated modification of models and code during evolution. Intermittent inconsistencies are a common occurrence during co-changing. A partial co-change is the period in which the developer changed, say, the model but has not yet propagated the change to the code. Inconsistency feedback can be provided to developers for helping them to complete partial co-changes. However, there is no evidence whether such inconsistency feedback is useful to developers. To investigate this problem, we conducted a controlled experiment with 36 subjects who were required to complete ten partially completed change tasks between models and code of two non-trivial systems [31]. The tasks were of different levels of complexity depending on how many model diagrams they affected. All subjects had to work on all change tasks but sometimes with and sometimes without inconsistency feedback. We then measured differences between task effort and correctness. We found that when subjects were given inconsistency feedback during tasks, they were 268% more likely to complete the co-change correctly compared to when they were not given inconsistency feedback. We also found that when subjects were not given inconsistency feedback, they nearly always failed in completing co-change tasks with high complexity where the partially completed changes were spread across different diagrams in the model. These findings suggest that inconsistency feedback (i.e. detection and repair) should form an integral part of co-changing, regardless of whether the code or the model changes first. Furthermore, these findings suggest that merely having access to changes (as with the given partially completed changes) is insufficient for effective co-changing.

6.4.3.2. Detecting and Exploring Side Effects when Repairing Model Inconsistencies

When software models change, developers often fail in keeping them consistent. Automated support in repairing inconsistencies is widely addressed. Yet, merely enumerating repairs for developers is not enough. A repair can as a side effect cause new unexpected inconsistencies (negative) or even fix other inconsistencies as well (positive). To make matters worse, repairing negative side effects can in turn cause further side effects. Current approaches do not detect and track such side effects in depth, which can increase developers’ effort and time spent in repairing inconsistencies. This work [66] presents an automated approach for detecting and tracking the consequences of repairs, i.e. side effects. It recursively explores in depth positive and negative side effects and identifies paths and cycles of repairs. This work further ranks repairs based on side effect knowledge so that developers may quickly find the relevant ones. Our approach and its tool implementation have been empirically assessed on 14 case studies from industry, academia, and GitHub. Results show that both positive and negative side effects occur frequently. A comparison with three versioned models showed the usefulness of our ranking strategy based on side effects. It showed that our approach’s top prioritized repairs are those that developers would indeed choose. A controlled experiment with 24 participants further highlights the significant influence of side effects and of our ranking of repairs on developers. Developers who received side effect knowledge chose far more repairs with positive side effects and far less with negative side effects, while being 12.3% faster, in contrast to developers who did not receive side effect knowledge.

6.4.3.3. Supporting A Flexible Grouping Mechanism for Collaborating Engineering Teams

Most engineering tools do not provide much support for collaborating teams and today's engineering knowledge repositories lack flexibility and are limited. Engineering teams have different needs and their team members have different preferences on how and when to collaborate. These needs may depend on the individual work style, the role an engineer has, and the tasks they have to perform within the collaborating group. However, individual collaboration is insufficient and engineers need to collaborate in groups. This work [65] presents a collaboration framework for collaborating groups capable of providing synchronous and asynchronous mode of collaboration. Additionally, our approach enables engineers to mix these collaboration modes to meet the preferences of individual group members. We evaluate the scalability of this framework using four real life large collaboration projects. These projects were found from GitHub and they were under active development by the time of evaluation. We have tested our approach creating groups of different sizes for each project. The results showed that our approach scales to support every case for the groups created. Additionally, we scouted the literature and discovered studies that support the usefulness of different groups with collaboration styles.

6.4.4. Software diversification

6.4.4.1. The Maven Dependency Graph: a Temporal Graph-based Representation of Maven Central

The Maven Central Repository provides an extraordinary source of data to understand complex architecture and evolution phenomena among Java applications. As of September 6, 2018, this repository includes 2.8M artifacts (compiled piece of code implemented in a JVM-based language), each of which is characterized with metadata such as exact version, date of upload and list of dependencies towards other artifacts. Today, one who wants to analyze the complete ecosystem of Maven artifacts and their dependencies faces two key challenges: (i) this is a huge data set; and (ii) dependency relationships among artifacts are not modeled explicitly and cannot be queried. In this work [55], we present the Maven Dependency Graph. This open source data set provides two contributions: a snapshot of the whole Maven Central taken on September 6, 2018, stored in a graph database in which we explicitly model all dependencies; an open source infrastructure to query this huge dataset.

6.4.4.2. The Emergence of Software Diversity in Maven Central

Maven artifacts are immutable: an artifact that is uploaded on Maven Central cannot be removed nor modified. The only way for developers to upgrade their library is to release a new version. Consequently, Maven Central accumulates all the versions of all the libraries that are published there, and applications that declare a dependency towards a library can pick any version. In this work [59], we hypothesize that the immutability of Maven artifacts and the ability to choose any version naturally support the emergence of software diversity within Maven Central. We analyze 1,487,956 artifacts that represent all the versions of 73,653 libraries. We observe that more than 30% of libraries have multiple versions that are actively used by latest artifacts. In the case of popular libraries, more than 50% of their versions are used. We also observe that more than 17% of libraries have several versions that are significantly more used than the other versions. Our results indicate that the immutability of artifacts in Maven Central does support a sustained level of diversity among versions of libraries in the repository.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. ADR Nokia

- Coordinator: Inria
- Dates: 2017-2021

- Abstract: The goal of this project is to integrate a chaos engineering principles to IoT Services frameworks to improve the robustness of the software-defined network services using this approach and to explore the concept of equivalence for software-defined network services and propose an approach to constantly evolve the attack surface of the network services.

7.1.2. BCOM

- Coordinator: UR1
- Dates: 2018-2024
- Abstract: The aim of the Falcon project is to investigate how to improve the resale of available resources in private clouds to third parties. In this context, the collaboration with DiverSE mainly aims at working on efficient techniques for the design of consumption models and resource consumption forecasting models. These models are then used as a knowledge base in a classical autonomous loop.

7.1.3. 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.

7.1.4. GLOSE Demonstrator

- Partners: Inria/Safran
- Dates: 2019-2020
- Abstract: Demonstrator illustrating the technologies involved in the WP5 off the GLOSE project. The use case chosen for the demonstrator is the high-level description of a remote control drone system, whose the main objective is to illustrate the design and simulation of the main functional chains, the possible interactivity with the model in order to raise the level of understanding over the models built, and possibly the exploration of the design space.

7.1.5. OneShotSoftware

- 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.

7.1.6. Kereval

- Partners: INSA Rennes/Kereval
- Dates: 2019-2022
- Abstract: Front-ends testing in a DevOps context, Romain Lebouc's PhD Cifre project.

7.1.7. Obeo

- Partners: Inria/Obéo
- Dates: 2017-2020
- Abstract: Web engineering for domain-specific modeling languages, Fabien Coulon's PhD Cifre project.

7.1.8. OKWind

- Partners: UR1/OKWind
- Dates: 2017-2020
- Abstract: Models@runtime to improve self-consumption of renewable energies, Alexandre Rio's PhD Cifre project.

7.1.9. Orange

- Partners: UR1/Orange
- Dates: 2016-2019
- Abstract: Modelling and evaluating security of authentication paths, Youssou Ndiaye's PhD Cifre project.

7.1.10. Keolis

- Partners: UR1/Keolis
- Dates: 2018-2021
- Abstract: Urban mobility: machine learning for building simulators using large amounts of data, Gauthier LYAN's PhD Cifre project.

7.1.11. FaberNovel

- Partners: UR1/FaberNovel
- Dates: 2018-2021
- Abstract: Abstractions for linked data and the programmable web, Antoine Cheron's PhD Cifre project.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. PEC – Pôle d'Excellence Cyber

- Coordinator: Université de Rennes 1
- Dates: 2016-2019
- Abstract: Formal and Executable Specification of domain-specific language families.

8.2. National Initiatives

8.2.1. ANR

8.2.1.1. 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 run time, 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).

8.2.2. DGA

8.2.2.1. LangComponent (CYBERDEFENSE)

- Coordinator: DGA
- Partners: DGA MI, Inria
- Dates: 2019-2022
- 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.

8.2.3. Cominlabs

8.2.3.1. 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) 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 workings, 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.

8.3. European Initiatives

8.3.1. FP7 & H2020 Projects

8.3.1.1. 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 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 raises 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.

8.3.2. 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 multi-objective elasticity in cloud management; model-driven development.

KTH, the Royal Institute of Technology (Sweden): continuous software testing, perturbation and diversification.

McGill University (Canada): language reuse, model composition, and models for sustainability.

CWI (The Netherlands): language engineering.

JKU Linz (Austria): model analysis and Model-Based DevOps.

RWTH Aachen (Germany): models for industry 4.0

8.4. International Initiatives

8.4.1. Inria International Labs

IIL CWI-Inria

Associate Team involved in the International Lab:

8.4.1.1. ALE

- Title: Agile Language Engineering
- International Partner (Institution - Laboratory - Researcher):
  - CWI (Netherlands) Tijs van der Storm
- 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).

8.4.2. Inria International Partners

8.4.2.1. Informal International Partners

- Université de Montréal (Canada)
- McGill University (Canada)
- University of Alabama (USA)
- University of Lancaster (UK)
- University of Namur (Belgium)
- Universita degli Studi di Cagliari (Italy)
- Università degli Studi dell'Aquila (Italy)
- JKU Linz (Austria)
- TU Wien (Austria)
- Michigan State University (MSU)
- RWTH Aachen University (Germany)
- KTH (Sweden)

8.4.3. Participation in Other International Programs

The GEMOC studio has been sustained through the creation of a Research Consortium at the Eclipse Foundation.

8.4.3.1. 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 provided 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 a GEMOC 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.

8.5. International Research Visitors

8.5.1. Visits of International Scientists

- Yves Le Traon, Professor at the University of Luxembourg, visited the team in June, July and October 2019.
- Nelly Bencomo, Lecturer in Computer Science Aston University, UK, visited the team from October 2019 to June 2020.
- Martin Montperrus, Professor at KTH, Sweden, visited the team in December 2019.
- Nicolas Harrand, PhD Student at KTH, Sweden, visited the team in December 2019.
- Paul Temple, postdoc at University de Namur, visited the team in February 2019.
- Thomas Degueule, postdoc at CWI, visited the team in December 2019
- Alfonso Pierantonio, Associate Professor at Università degli Studi dell'Aquila, visited the team in June 2019
- Mark van den Brand, Professor at Eindhoven University of Technology, visited the team in June 2019

8.5.2. Visits to International Teams

- Pierre JeanJean visited CWI for 1 week in December 2019 in the context of the Associated Team ALE.
- Benoit Combemale made several short visits at CWI in the context of the Associated Team ALE, visited McGill University in June 2019, and visited TU Eindhoven in November 2019.
- Olivier Barais made several short visits at KTH in the context of a collaboration with Prof Monperrus and Prof Baudry.
- Djamel E. Khelladi made a one week research visit in December 2019 to the DIRO laboratory at University of Montreal, Canada.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Organisation

9.1.1.1. General Chair, Scientific Chair

- Mathieu Acher was the program committee co-chair of the 14th International Working Conference on Variability Modelling of Software-Intensive Systems (VaMoS)

9.1.1.2. Member of the Organizing Committees

Olivier Barais was the financial chair of the IEEE Mascots conference.

9.1.2. Scientific Events: Selection

9.1.2.1. Chair of Conference Program Committees

- Mathieu Acher organized the Fourth international workshop on software product line teaching (SPLTea 2019).
- Mathieu Acher organized the First international workshop on languages for modelling variability (MODEVAR 2019).
- Mathieu Acher organized the Seventh international workshop on reverse variability engineering (REVE 2019).
- Mathieu Acher organized the 3rd ACM SIGSOFT International Workshop on Machine Learning Techniques for Software Quality Evaluation, MaLTeSQuE@ESEC/SIGSOFT FSE 2019.

9.1.2.2. Member of the Conference Program Committees

Jean-Marc Jézéquel:

- PC member for SPLC 2019 Industry Track
- PC member for SEAMS 2019
- PC member for ICSE 2020

Olivier Barais:

- PC member for the 17th International Conference on Software Engineering and Formal Methods (SEFM 2019)
- PC Member for CIEL 2019
- PC Member for VaMoS 2019
- PC Member for VoSE 2019
- PC Member for SAC 2019, SATTA track - Software Architecture: Theory, Technology, and Applications
- PC Member for CCgrid 2020

Djamel E. Khelladi:

- PC member for Poster track at MODELS 2019
- PC member for REVE workshop 2019
- PC member for Models and Evolution (ME) Workshop 2019
- PC member for WETICE 2019 Validating Software for Critical Systems Track

Arnaud Blouin:

- PC member for SAC'2020
- PC member for PACM EICS Q4 2019
- PC member for Workshop on Human Factors in Modeling at MODELS'2019 (HuFaMo), 2019

Mathieu Acher:

- PC member for SPLC 2019 Research Track
- PC member for ICSE SEIP 2020
- PC member for VaMoS 2019

Benoit Combemale:

- PC chair (Foundation track) for ECMFA 2019
- PC member for ICMT'19
- Selection Committee member for the Doctoral Symposium at MODELS'19
- PC member for the MLE'19 workshop at MODELS'19
- PC member for the DevOps@MODELS'19 workshop at MODELS'19
- PC member for the FlexMDE'19 workshop at MODELS'19
- PC member for the PNSE'19 workshop

9.1.2.3. Reviewer

Olivier Barais: MASCOTS 2019, ICSE 2019, ASE 2019
Arnaud Blouin: ICSE 2019, ISSTA 2019
Johann Bourcier: SEAA 2019
Mathieu Acher: TSE 2019, JSS 2019, ICSE 2019

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

Jean-Marc Jézéquel:

- Associate Editor in Chief of SOSYM
- Associate Editor in Chief of IEEE Computer
- Associate Editor of JSS
- Associate Editor of JOT

Benoit Combemale:

- Deputy Editor-in-Chief of JOT
- Member of the Editorial Board of the Springer Journal on Software and Systems Modeling (SoSyM)
- Member of the Editorial Board of the Elsevier Journal of Computer Languages (COLA)
- Member of the Editorial Board of the Springer Journal of Software Quality
- Member of the Advisory Board of the Elsevier Journal on Science of Computer Programming (SCP, Software Section)

9.1.3.2. Reviewer - Reviewing Activities

Olivier Barais:

- Journal of Software and System Modeling (SoSyM)

Djamel E. Khelladi:

- Journal of Software and System Modeling (SoSyM)

Arnaud Blouin:

- Journal of Software and System Modeling (SoSyM)
- Journal of Visual Languages and Computing (JVLC)
- Proceedings of the ACM for EICS 2019 and 2020 (PACM-EICS)

Mathieu Acher:

- Transactions on Software Engineering (TSE)
- Journal of Systems and Software (JSS)
- Editor with Myra B. Cohen of the Special issue on systems and software product line engineering. Journal of Systems and Software (JSS) 154: 110-111 (2019).

Benoit Combemale:

- Journal of Software and System Modeling (SoSyM)
- Journal of Empirical Software Engineering (ESE)

9.1.4. Invited Talks

Jean-Marc Jézéquel:

- *Keynote speech Models@Runtime workshop associated to MODELS2019*

Djamel E. Khelladi:

- Talk at the *IPA Fall Days* in Wageningen, Netherlands.

Johann Bourcier:

- Talk at the UBS university in Vannes, France.

Benoit Combemale:

- *Smart Modeling: On the Convergence of Scientific and Engineering Models*, Invited talk at Lancaster university, UK (14/11/19).
- *Bringing Intelligence to Sociotechnical IoT Systems: Modeling Opportunities and Challenges*, Keynote at MDDE4IoT'19, DE (15/09/19).
- *Breathe Life Into Your IDE*, Talk at LangDev'19, NL (22/03/19).

9.1.5. Leadership within the Scientific Community

Jean-Marc Jézéquel is an elected member of the Board of *InformaticsEurope*.

Benoit Combemale:

- Chair of the Steering committee of the ACM SIGPLAN conference SLE
- Founding member and member of the advisory board of the GEMOC initiative.
- Chair of the Eclipse Research Consortium GEMOC and the Eclipse Project GEMOC Studio.

Mathieu Acher:

- Member of the Steering committee of SPLC conference

Arnaud Blouin: founding member and co-organiser of the French GDR-GPL research action on Software Engineering and Human-Computer Interaction (GL-IHM).

9.1.6. Scientific Expertise

Olivier Barais is an external expert for the H2020 ENACT project. Olivier Barais provides scientific expertises for DGRI international program (20 proposals per year).

9.1.7. Research Administration

Jean-Marc Jézéquel is Director of IRISA (UMR 6074). He is Coordinator of the academic club of the French Cyber-defense Excellence Cluster, and Director of the Rennes Node of EIT Digital.

9.2. Teaching - Supervision - Juries

9.2.1. 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, and Mathieu Acher teaches about 200h in these domains, and Benoit Combemale teaching about 100h, for a grand total of about 1300 hours, including several courses at ENSTB, Supelec, ENS Rennes and ENSAI Rennes engineering school.

Olivier Barais is deputy director of the electronics and computer science teaching department of the University of Rennes 1. Olivier Barais is the head of the final year of the Master in Computer Science at the University of Rennes 1. Johann Bourcier is the head of the Information Technology department and member of the management board at the ESIR engineering school in Rennes. Arnaud Blouin is in charge of industrial relationships for the computer science department at INSA Rennes and elected member of this CS department council.

The DIVERSE team also hosts several MSc and summer trainees every year.

Mathieu Acher gave courses to EJCP (Ecole Jeune Chercheur en Programmation), a well-known, national training school for PhD students http://ejcp2019.icube.unistra.fr/

9.2.2. Supervision

- PhD in progress: Alejandro Gomez Boix, *Distributed counter-measure against browser fingerprinting*, 2016, B. Baudry, D. Bromberg.
- PhD in progress: Jean-Émile Dartois, *Efficient resources management for hybrid cloud computing*, 2016, O. Barais, Jalil Boukhobza.
- PhD in progress: Alexandre Rio, *Demand Side Management A model driven approach to promote energy self-consumption* , 2016, O. Barais, Y. Morel
- PhD in progress: Dorian Leroy, *A generic and generative white-box testing framework for model transformations*, 2017, B. Combemale.
- PhD in progress: Fabien Coulon, *Web engineering for domain-specific modeling languages*, 2017, B. Combemale, S. Begaudeau.
- PhD in progress: June Benvegnu-Sallou, Decision support for the assessment of risks associated with the operation of underground environments, 2018, J-R. De dreuzy, B. Combemale, J. Bourcier.
- PhD in progress: Pierre JeanJean, *Refining simulators by analyzing execution traces of complex systems*, 2018, O. Barais, B. Combemale.
- PhD in progress: Alif Akbar-pranata, *Chaos Engineering for IoT and Network Services*, 2018, O. Barais, J. Bourcier.
- PhD in progress: Antoine Cheron, *Abstractions for linked data and the programmable web*, 2018, O. Barais, J. Bourcier.
- PhD in progress: Gauthier Lyan, *Urban mobility: machine learning for building simulators using large amounts of data*, 2018, J-M. Jééquel, D. Gross Amblard.
- PhD in progress: Hugo Martin, *Learning variability*, 2018, M. Acher.
- PhD in progress: Luc Lesoil, *Deep variability in large-scaled systems*, 2019, M. Acher. A. Blouin, JM Jézéquel.
- PhD in progress: Emmanuel Chebbi , *Domain-Specific Language Reuse*, 2019, B. Combemale, Olivier Barais, G. Leguernic
- PhD in progress: Mohamed Handaoui, *Scheduling Big Data applications in the cloud: the case of machine learning algorithms*, 2019, Jalil Boukhobza. Olivier Barais.
- PhD in progress: Romain Lebouc, *Testing front-ends in a DevOps context*, 2019, Arnaud Blouin, Noël Plouzeau, Alain Ribault.
- PhD : Manuel Leduc, *Formal and Executable Specification of domain-specific language families*, dec 2019, O. Barais, B. Combemale, G. Leguernic
- PhD : Youssou NDiaye, *Modelling and evaluating security of authentication paths*, dec 2019, N. Aillery, O. Barais, A. Blouin, A. Bouabdallah
- PhD: Ludovic Mouline, *Omniscient fault localization in IOT systems using model-driver data analytics*, nov 2019, O. Barais, Y. Le-Traon, J. Bourcier [23]
- PhD: Oscar Luis, *Automatic test amplification*, dec 2019, B. Baudry
- PhD: Manuel Leduc, *On modularity and performances of external domain-specific language implementations*, dec 2019, B. Combemale, O. Barais
- HdR : Arnaud Blouin, Contribution to the Engineering of User Interfaces, Univ Rennes, 30/08/2019 [22]

9.2.3. Juries

9.2.3.1. Jean-Marc Jézéquel

was in the examination committee of the following PhD and HDR thesis:

- Yoann Blein, April 2019, Univ Grenobles Alpes (Examiner)
- Jorge Roda, December 2019, Univ Sevilla (Examiner)
- Emilien Lavigne, December 2019, Univ Bretagne Occidentale (President)

9.2.3.2. Olivier Barais

was in the examination committee of the following PhD and HDR thesis:

- Benjamen Benni, December 2019, Univ Nice (Reviewer)
- Ludovic Mouline, November 2019, Univ Luxembourg (Supervisor)
- Leduc Manuel, December 2019, Univ Rennes (Supervisor)
- Ndiaye Youssou, December 2019, Univ Rennes (Supervisor)
- Davide Frey, December 2019, Univ Rennes (President)
- Hanyang CAO, February 2019, Univ Bordeaux (Reviewer)

9.2.3.3. Arnaud Blouin

was in the examination committee of the following PhD thesis:

- Youssou Ndiaye, December 2019, Univ Rennes (Supervisor)

9.2.3.4. Mathieu Acher

was in the examination committee of the following PhD thesis:

- Frederic Verdier, December 2019, Univ Montpellier (Examiner)

9.2.3.5. Johann Bourcier

was in the examination committee of the following PhD thesis:

- Ludovic Mouline, November 2019, Univ Luxembourg (Supervisor)

9.2.3.6. Benoit Combemale

was in the examination committee of the following PhD thesis:

- Théo Zimmermann, Univ Paris (December 2019), (Reviewer)
- Manuel Leduc, Univ Rennes 1 (December 2019), (Supervisor)

9.3. Popularization

9.3.1. Interventions

- STAMP - Orange Test and Dev Day, Grenoble, France - Nov 2019
- STAMP - Eastern Paris Dev Meetup, Lunatech, Chessy, France - Jul 2019
- STAMP Breakout Session & Talks - OW2con'19, Orange Gardens, Paris Châtillon - Jun 2019
- STAMP Workshop for EC-DGIT (European Commission DGIT) - Brussels - May 2019
- STAMP - Devoxx Paris - Apr 2019
- STAMP & CI testing automation - BreizhCamp 2019 - Mar 2019
- STAMP Talk and Workshop - Station-F, Paris - Feb 2019
- Tech talk at Alten - Rennes - Dec 2019
- Tech talk at Harmonic - Cesson-Sévigné - Oct 2019
- Tech talk at Veonum - Rennes - Sep 2019
- Tech talk at Nokia - Rennes - Jun 2019
- Tech talk at OrangeLabs : Hands-on Workshop - Cesson-Sévigné - May 2019
- Tech talk at Solocal and Kereval - Rennes - Mar 2019
- Tech talk at FaberNovel - Test amplification - Mar 2019
- Sciences de l'Ingénieur au féminin - Lycée Sévigné, Cesson-Sévigné - Nov 2019
- *J'peux pas j'ai informatique* - Inria, Rennes - Apr 2019

- *Petit-déjeuenr des métiers* - Lycée Sévigné, Cesson-Sévigné - Mar 2019 (je ne suis pas sure de la date)
- Devoxx 2019 - Les APIs hypermedia expliquées simplement (17-19 avril 2019) ("While web pages are full of hyperlinks and hypermedia, web APIs are not. However, augmenting web APIs with links is at the core of the REST architectural style, but only 5% of APIs actually do it. That's right, only 5%, despite most people saying that they provide RESTful APIs. In this talk, we discuss what an hypermedia-powered API looks like, what problems of distributed systems integration and composition it can solve and what it would change in frontend code to have such APIs.
- Embedded Linux conference Europe [54] – Given a configuration, can humans know in advance the size, the compilation time, or the boot time of a Linux kernel? Owing to the huge complexity of Linux (there are more than 15000 options with hard constraints and subtle interactions), machines should rather assist contributors and integrators in mastering the configuration space of the kernel. In this talk, we introduce TuxML an OSS tool based on Docker/Python to massively gather data about thousands of kernel configurations. Mathieu will describe how 200K+ configurations have been automatically built and how machine learning can exploit this information to predict properties of unseen Linux configurations, with different use cases (identification of influential/buggy options, finding of small kernels, etc.) The vision is that a continuous understanding of the configuration space is undoubtedly beneficial for the Linux community, yet several technical challenges remain in terms of infrastructure and automation.
- API Days - Backend is the new frontend (9-11 décembre 2019)

10. Bibliography

Major publications by the team in recent years

[1] 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

[2] 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^{o} 1, p. 16:1–16:26, https://hal.inria.fr/hal-01182103

[3] A. BLOUIN, V. LELLI, B. BAUDRY, F. COULON.*User Interface Design Smell: Automatic Detection and Refactoring of Blob Listeners*, in "Information and Software Technology", May 2018, vol. 102, p. 49-64 [*DOI :* 10.1016/J.INFSOF.2018.05.005], https://hal.inria.fr/hal-01499106

[4] M. BOUSSAA, O. BARAIS, G. SUNYÉ, B. BAUDRY.*Leveraging metamorphic testing to automatically detect inconsistencies in code generator families*, in "Software Testing, Verification and Reliability", December 2019 [*DOI :* 10.1002/STVR.1721], https://hal.inria.fr/hal-02422437

[5] E. BOUSSE, D. LEROY, B. COMBEMALE, M. WIMMER, B. BAUDRY.*Omniscient Debugging for Executable DSLs*, in "Journal of Systems and Software", March 2018, vol. 137, p. 261-288 [*DOI :* 10.1016/J.JSS.2017.11.025], https://hal.inria.fr/hal-01662336

[6] 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^{o} 4, p. 1794–1841 [*DOI :* 10.1007/S10664-014-9357-1], https://hal.inria.fr/hal-01096969

[7] 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

[8] 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^{o} 3, p. 72-86 [*DOI :* 10.1515/POPETS-2017-0029], https://hal.archives-ouvertes.fr/hal-01611048

[9] J.-E. DARTOIS, J. BOUKHOBZA, A. KNEFATI, O. BARAIS.*Investigating Machine Learning Algorithms for Modeling SSD I/O Performance for Container-based Virtualization*, in "IEEE transactions on cloud computing", 2019, vol. 14, p. 1-14 [*DOI :* 10.1109/TCC.2019.2898192], https://hal.inria.fr/hal-02013421

[10] 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

[11] 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

[12] J. A. GALINDO DUARTE, 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

[13] 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

[14] A. HALIN, A. NUTTINCK, M. ACHER, X. DEVROEY, G. PERROUIN, B. BAUDRY.*Test them all, is it worth it? Assessing configuration sampling on the JHipster Web development stack*, in "Empirical Software Engineering", July 2018, p. 1–44 [*DOI :* 10.1007/S10664-018-9635-4], https://hal.inria.fr/hal-01829928

[15] 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^{o} 2, p. 905-920, https://hal.inria.fr/hal-00829839

[16] 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

[17] M. LEDUC, T. DEGUEULE, E. VAN WYK, B. COMBEMALE.*The Software Language Extension Problem*, in "Software and Systems Modeling", 2019, p. 1-4, https://hal.inria.fr/hal-02399166

[18] 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

[19] P. TEMPLE, M. ACHER, J.-M. JEZEQUEL, O. BARAIS.*Learning-Contextual Variability Models*, in "IEEE Software", November 2017, vol. 34, n^o 6, p. 64-70 [*DOI :* 10.1109/MS.2017.4121211], https://hal.inria.fr/hal-01659137

[20] P. TEMPLE, M. ACHER, J.-M. JÉZÉQUEL.*Empirical Assessment of Multimorphic Testing*, in "IEEE Transactions on Software Engineering", July 2019, p. 1-21 [*DOI :* 10.1109/TSE.2019.2926971], https://hal.inria.fr/hal-02177158

[21] O. L. VERA-PÉREZ, B. DANGLOT, M. MONPERRUS, B. BAUDRY.*A Comprehensive Study of Pseudo-tested Methods*, in "Empirical Software Engineering", 2018, p. 1-33 [*DOI :* 10.1007/S10664-018-9653-2], https://hal.inria.fr/hal-01867423

Publications of the year

Doctoral Dissertations and Habilitation Theses

[22] A. BLOUIN.*Contribution to the Engineering of User Interfaces*, Université de Rennes 1 [UR1], August 2019, Habilitation à diriger des recherches, https://tel.archives-ouvertes.fr/tel-02354530

[23] L. MOULINE.*Towards a Modelling Framework with Temporal and Uncertain Data for Adaptive Systems*, Université Rennes 1 ; Université du Luxembourg, November 2019, https://hal.inria.fr/tel-02404304

Articles in International Peer-Reviewed Journal

[24] 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 2019, vol. 27, n^o 1, p. 307-347 [*DOI :* 10.1007/S11219-017-9400-8], https://hal.inria.fr/hal-01688247

[25] M. BOUSSAA, O. BARAIS, G. SUNYÉ, B. BAUDRY.*Leveraging metamorphic testing to automatically detect inconsistencies in code generator families*, in "Software Testing, Verification and Reliability", December 2019 [*DOI :* 10.1002/STVR.1721], https://hal.inria.fr/hal-02422437

[26] 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", February 2019, p. 1–37 [*DOI :* 10.1007/S10270-017-0598-5], https://hal.inria.fr/hal-01614377

[27] J.-M. BRUEL, B. COMBEMALE, E. M. GUERRA, J.-M. JÉZÉQUEL, J. KIENZLE, J. DE LARA, G. MUSSBACHER, E. SYRIANI, H. VANGHELUWE.*Comparing and Classifying Model Transformation Reuse Approaches across Metamodels*, in "Software and Systems Modeling", 2019, p. 1-22, https://hal.inria.fr/hal-02317864

[28] B. DANGLOT, O. L. VERA-PÉREZ, B. BAUDRY, M. MONPERRUS.*Automatic Test Improvement with DSpot: a Study with Ten Mature Open-Source Projects*, in "Empirical Software Engineering", 2019, p. 1-35 [*DOI :* 10.1007/S10664-019-09692-Y], https://hal.inria.fr/hal-01923575

[29] B. DANGLOT, O. L. VERA-PÉREZ, Z. YU, A. ZAIDMAN, M. MONPERRUS, B. BAUDRY.*A Snowballing Literature Study on Test Amplification*, in "Journal of Systems and Software", August 2019, vol. 157, p. 1-16 [*DOI :* 10.1016/J.JSS.2019.110398], https://hal.inria.fr/hal-02290742

[30] J.-E. DARTOIS, J. BOUKHOBZA, A. KNEFATI, O. BARAIS.*Investigating Machine Learning Algorithms for Modeling SSD I/O Performance for Container-based Virtualization*, in "IEEE transactions on cloud computing", 2019, vol. 14, p. 1-14 [*DOI :* 10.1109/TCC.2019.2898192], https://hal.inria.fr/hal-02013421

[31] G. KANAKIS, D. E. KHELLADI, S. FISCHER, M. TRÖLS, A. EGYED.*An Empirical Study on the Impact of Inconsistency Feedback during Model and Code Co-changing*, in "The Journal of Object Technology", 2019, vol. 18, n^{o} 2, p. 10:1-21 [*DOI :* 10.5381/JOT.2019.18.2.A10], https://hal.inria.fr/hal-02192486

[32] J. KIENZLE, G. MUSSBACHER, B. COMBEMALE, L. BASTIN, N. BENCOMO, J.-M. BRUEL, C. BECKER, S. BETZ, R. CHITCHYAN, B. CHENG, S. KLINGERT, R. PAIGE, B. PENZENSTADLER, N. SEYFF, E. SYRIANI, C. C. VENTERS.*Towards Model-Driven Sustainability Evaluation*, in "Communications of the ACM", 2019, p. 1-10, https://hal.inria.fr/hal-02146543

[33] J. KIENZLE, G. MUSSBACHER, B. COMBEMALE, J. DEANTONI.*A Unifying Framework for Homogeneous Model Composition*, in "Software and Systems Modeling", January 2019, p. 1-19 [*DOI :* 10.1007/S10270-018-00707-8], https://hal.inria.fr/hal-01949050

[34] M. LEDUC, T. DEGUEULE, E. VAN WYK, B. COMBEMALE.*The Software Language Extension Problem*, in "Software and Systems Modeling", 2019, p. 1-4, https://hal.inria.fr/hal-02399166

[35] B. LELANDAIS, M.-P. OUDOT, B. COMBEMALE.*Applying Model-Driven Engineering to High-Performance Computing: Experience Report, Lessons Learned, and Remaining Challenges*, in "Computer Languages, Systems and Structures", 2019, p. 1-19, https://hal.inria.fr/hal-02296030

[36] A. PIERANTONIO, M. VAN DEN BRAND, B. COMBEMALE.*The JOT Journal: Towards a Rising Generation*, in "Journal of Object Technology (JOT)", 2019, vol. 18, n^{o} 1, p. 1-3 [*DOI :* 10.5381/JOT.2019.18.1.E1], https://hal.inria.fr/hal-02408017

[37] P. TEMPLE, M. ACHER, J.-M. JÉZÉQUEL.*Empirical Assessment of Multimorphic Testing*, in "IEEE Transactions on Software Engineering", July 2019, p. 1-21, forthcoming [*DOI :* 10.1109/TSE.2019.2926971], https://hal.inria.fr/hal-02177158

[38] A. WORTMANN, O. BARAIS, B. COMBEMALE, M. WIMMER.*Modeling Languages in Industry 4.0: An Extended Systematic Mapping Study*, in "Software and Systems Modeling", 2019, p. 1-28 [*DOI :* 10.1007/S10270-019-00757-6], https://hal.inria.fr/hal-02282028

Invited Conferences

[39] B. COMBEMALE.*Bringing Intelligence to Sociotechnical IoT Systems: Modeling Opportunities and Challenges*, in "MDE4IoT 2019 - 3nd International Workshop on Model-Driven Engineering for the Internet-of-Things", Munich, Germany, September 2019, p. 1-2, https://hal.inria.fr/hal-02285737

International Conferences with Proceedings

[40] M. ACHER, T. ZIADI, R. E. LOPEZ-HERREJON, J. MARTINEZ.*Seventh international workshop on reverse variability engineering (REVE 2019)*, in "SPLC 2019 - 23rd International Systems and Software Product Line Conference", Paris, France, ACM Press, September 2019, 1, https://hal.archives-ouvertes.fr/hal-02268373

[41] B. AMAND, M. CORDY, P. HEYMANS, M. ACHER, P. TEMPLE, J.-M. JÉZÉQUEL.*Towards Learning-Aided Configuration in 3D Printing: Feasibility Study and Application to Defect Prediction*, in "VAMOS 2019 - 13th

International Workshop on Variability Modelling of Software-Intensive Systems", Leuven, Belgium, ACM, February 2019, p. 1-9 [*DOI :* 10.1145/3302333.3302338], https://hal.inria.fr/hal-01990767

[42] M. BAGHERZADEH, K. JAHED, B. COMBEMALE, J. DINGEL.*Live-UMLRT: A Tool for Live Modeling of UML-RT Models*, in "MODELS 2019 - ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems", Munich, Germany, IEEE, September 2019, p. 743-747 [*DOI :* 10.1109/MODELS-C.2019.00115], https://hal.inria.fr/hal-02407932

[43] J.-E. DARTOIS, J. BOUKHOBZA, V. FRANCOISE, O. BARAIS.*Tracking Application Fingerprint in a Trustless Cloud Environment for Sabotage Detection*, in "MASCOTS 2019 - 27th IEEE International Symposium on the Modeling, Analysis, and Simulation of Computer and Telecommunication Systems", Rennes, France, IEEE, October 2019, p. 74-82 [*DOI :* 10.1109/MASCOTS.2019.00018], https://hal.archives-ouvertes.fr/hal-02303153

[44] J.-E. DARTOIS, I. MERIAU, M. HANDAOUI, J. BOUKHOBZA, O. BARAIS.*Leveraging cloud unused resources for Big data application while achieving SLA*, in "MASCOTS 2019 - 27th IEEE International Symposium on the Modeling, Analysis, and Simulation of Computer and Telecommunication Systems", Rennes, France, IEEE, October 2019, p. 1-2, https://hal.inria.fr/hal-02362257

[45] A. GÓMEZ-BOIX, D. FREY, Y.-D. BROMBERG, B. BAUDRY.*A Collaborative Strategy for mitigating Tracking through Browser Fingerprinting*, in "MTD 2019 - 6th ACM Workshop on Moving Target Defense", London, United Kingdom, November 2019, p. 1-12 [*DOI :* 10.1145/3338468.3356828], https://hal.inria.fr/hal-02282591

[46] M. HEINZ, R. LÄMMEL, M. ACHER.*Discovering Indicators for Classifying Wikipedia Articles in a Domain: A Case Study on Software Languages*, in "SEKE 2019 - The 31st International Conference on Software Engineering and Knowledge Engineering", Lisbonne, Portugal, July 2019, p. 1-6, https://hal.inria.fr/hal-02129131

[47] P. JEANJEAN, B. COMBEMALE, O. BARAIS.*From DSL Specification to Interactive Computer Programming Environment*, in "SLE 2019 - 12th ACM SIGPLAN International Conference on Software Language Engineering", Athènes, Greece, ACM, October 2019, p. 167-178 [*DOI :* 10.1145/3357766.3359540], https://hal.inria.fr/hal-02307953

[48] G. LE GUERNIC.*Experience Report on the Development of a Specialized Multi-view Multi-stakeholder Model-Based Engineering Framework*, in "DSM 2019 - 17th ACM SIGPLAN International Workshop on Domain-Specific Modeling", Athens, Greece, ACM Press, October 2019, p. 50-59 [*DOI :* 10.1145/3358501.3361237], https://hal.inria.fr/hal-02398053

[49] H. MARTIN, J. A. PEREIRA, P. TEMPLE, M. ACHER.*Machine Learning and Configurable Systems: A Gentle Introduction*, in "SPLC 2019 - 23rd International Systems and Software Product Line Conference", Paris, France, ACM, September 2019, p. 83-88 [*DOI :* 10.1145/3336294.3342383], https://hal.inria.fr/hal-02287459

[50] Y. NDIAYE, O. BARAIS, A. BLOUIN, A. BOUABDALLAH, N. AILLERY.*Requirements for preventing logic flaws in the authentication procedure of web applications*, in "SAC 2019 - 34th ACM/SIGAPP Symposium On Applied Computing", Limassol, Cyprus, April 2019, p. 1-9 [*DOI :* 10.1145/3297280.3297438], https://hal.inria.fr/hal-02087663

[51] Q. PLAZAR, M. ACHER, G. PERROUIN, X. DEVROEY, M. CORDY.*Uniform Sampling of SAT Solutions for Configurable Systems: Are We There Yet?*, in "ICST 2019 - 12th IEEE International Conference on Software Testing, Verification, and Validation", Xian, China, IEEE, April 2019, p. 240-251 [*DOI :* 10.1109/ICST.2019.00032], https://hal.inria.fr/hal-01991857

[52] M. RODRIGUEZ-CANCIO, B. COMBEMALE, B. BAUDRY.*Approximate Loop Unrolling*, in "CF 2019 - ACM International Conference on Computing Frontiers", Alghero, Sardinia, Italy, ACM, 2019, p. 94-105 [*DOI :* 10.1145/3310273.3323841], https://hal.inria.fr/hal-02407868

[53] P. TEMPLE, M. ACHER, G. PERROUIN, B. BIGGIO, J.-M. JÉZÉQUEL, F. ROLI.*Towards Quality Assurance of Software Product Lines with Adversarial Configurations*, in "SPLC 2019 - 23rd International Systems and Software Product Line Conference", Paris, France, ACM, September 2019, p. 277-288 [*DOI :* 10.1145/3336294.3336309], https://hal.inria.fr/hal-02287616

Conferences without Proceedings

[54] M. ACHER.*Learning the Linux Kernel Configuration Space: Results and Challenges*, in "ELC Europe 2019 - Embedded Linux Conference Europe 2019", Lyon, France, October 2019, p. 1-49, https://hal.inria.fr/hal-02342130

[55] *Best Paper*
A. BENELALLAM, N. HARRAND, C. SOTO-VALERO, B. BAUDRY, O. BARAIS.*The Maven Dependency Graph: a Temporal Graph-based Representation of Maven Central*, in "MSR 2019 - 16th International Conference on Mining Software Repositories", Montreal, Canada, ACM, May 2019, p. 344-348 [*DOI :* 10.1109/MSR.2019.00060], https://hal.archives-ouvertes.fr/hal-02080243.

[56] A. CHERON, J. BOURCIER, O. BARAIS, A. MICHEL.*Comparison Matrices of Semantic RESTful APIs Technologies*, in "ICWE 2019 - 19th International Conference On Web Engineering", Daejeon, South Korea, LNCS, Springer, June 2019, vol. 11496, p. 425-440 [*DOI :* 10.1007/978-3-030-19274-7_30], https://hal.archives-ouvertes.fr/hal-02114296

[57] J.-E. DARTOIS, H. B. RIBEIRO, J. BOUKHOBZA, O. BARAIS.*Cuckoo: Opportunistic MapReduce on Ephemeral and Heterogeneous Cloud Resources*, in "CLOUD 2019 - IEEE 12th International Conference on Cloud Computing", Milan, Italy, IEEE, July 2019, p. 1-8 [*DOI :* 10.1109/CLOUD.2019.00070], https://hal.archives-ouvertes.fr/hal-02179453

[58] A. RIO, Y. MAUREL, Y. BUGNI, O. BARAIS.*Benefits of Energy Management Systems on local energy efficiency, an agricultural case study*, in "SmartGridComm 2019 - IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids", Beijing, China, IEEE, October 2019, p. 1-7, https://hal.archives-ouvertes.fr/hal-02315327

[59] C. SOTO-VALERO, A. BENELALLAM, N. HARRAND, O. BARAIS, B. BAUDRY.*The Emergence of Software Diversity in Maven Central*, in "MSR 2019 - 16th International Conference on Mining Software Repositories", Montreal, Canada, ACM, May 2019, p. 333-343 [*DOI :* 10.1109/MSR.2019.00059], https://hal.archives-ouvertes.fr/hal-02080248

Scientific Books (or Scientific Book chapters)

[60] B. COMBEMALE, M. WIMMER.*Towards a Model-Based DevOps for Cyber-Physical Systems*, in "Software Engineering Aspects of Continuous Development", Springer-Verlag, 2019, p. 1-11, https://hal.inria.fr/hal-02407886

Research Reports

[61] M. ACHER, H. MARTIN, J. ALVES PEREIRA, A. BLOUIN, D. EDDINE KHELLADI, J.-M. JÉZÉQUEL.*Learning From Thousands of Build Failures of Linux Kernel Configurations*, Inria ; IRISA, June 2019, p. 1-12, https://hal.inria.fr/hal-02147012

[62] M. ACHER, H. MARTIN, J. A. PEREIRA, A. BLOUIN, J.-M. JÉZÉQUEL, D. E. KHELLADI, L. LESOIL, O. BARAIS.*Learning Very Large Configuration Spaces: What Matters for Linux Kernel Sizes*, Inria Rennes - Bretagne Atlantique, October 2019, https://hal.inria.fr/hal-02314830

[63] J. ALVES PEREIRA, H. MARTIN, M. ACHER, J.-M. JÉZÉQUEL, G. BOTTERWECK, A. VENTRESQUE.*Learning Software Configuration Spaces: A Systematic Literature Review*, Univ Rennes, Inria, CNRS, IRISA, June 2019, n^{o} 1-44, https://arxiv.org/abs/1906.03018 , https://hal.inria.fr/hal-02148791

[64] G. LE GUERNIC.*Experience Report on the Development of a Specialized Multi-view Multi-stakeholder Model-Based Engineering Framework (extended version)*, Inria Rennes - Bretagne Atlantique ; IRISA, December 2019, n^{o} RR-9283, https://hal.inria.fr/hal-02398051

Scientific Popularization

[65] G. KANAKIS, S. FISCHER, D. E. KHELLADI, A. EGYED.*Supporting A Flexible Grouping Mechanism for Collaborating Engineering Teams*, in "ICGSE 2019 - 14th ACM/IEEE International Conference on Global Software Engineering", Montreal, QC, Canada, IEEE, May 2019, p. 129-138 [*DOI :* 10.1109/ICGSE.2019.00033], https://hal.inria.fr/hal-02192482

[66] D. E. KHELLADI, R. KRETSCHMER, A. EGYED.*Detecting and Exploring Side Effects when Repairing Model Inconsistencies*, in "SLE 2019 - 12th ACM SIGPLAN International Conference on Software Language Engineering", Athènes, Greece, ACM, October 2019, p. 103-126 [*DOI :* 10.1145/3357766.3359546], https://hal.inria.fr/hal-02326034

Other Publications

[67] J. ALVES PEREIRA, M. ACHER, H. MARTIN, J.-M. JÉZÉQUEL.*Sampling Effect on Performance Prediction of Configurable Systems: A Case Study*, November 2019, working paper or preprint, https://hal.inria.fr/hal-02356290

[68] M. LEDUC, G. JOUNEAUX, T. DEGUEULE, G. LE GUERNIC, O. BARAIS, B. COMBEMALE.*Automatic generation of Truffle-based interpreters for Domain-Specific Languages*, December 2019, working paper or preprint, https://hal.inria.fr/hal-02395867

References in notes

[69] 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

[70] A. AVIZIENIS.*The N-version approach to fault-tolerant software*, in "Software Engineering, IEEE Transactions on", 1985, n° 12, p. 1491–1501

[71] F. BACHMANN, L. BASS.*Managing variability in software architectures*, in "SIGSOFT Softw. Eng. Notes", 2001, vol. 26, n° 3, p. 126–132

[72] 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° 4, p. 45–52

[73] 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

[74] 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° 1, p. 3–40

[75] 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

[76] 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° 1, p. 3–22

[77] 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

[78] A. BEUGNARD, J.-M. JÉZÉQUEL, N. PLOUZEAU.*Contract Aware Components, 10 years after*, in "WCSI", 2010, p. 1-11

[79] 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

[80] 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

[81] 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° 4, p. 387–404

[82] 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

[83] 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

[84] 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

[85] J. COPLIEN, D. HOFFMAN, D. WEISS.*Commonality and Variability in Software Engineering*, in "IEEE Software", 1998, vol. 15, n^o 6, p. 37–45

[86] 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^o 5, p. 593–615

[87] K. CZARNECKI, U. W. EISENECKER.*Generative programming: methods, tools, and applications*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 2000

[88] R. DEMILLI, A. J. OFFUTT.*Constraint-based automatic test data generation*, in "Software Engineering, IEEE Transactions on", 1991, vol. 17, n^o 9, p. 900–910

[89] 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^o 2, p. 182–197

[90] 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

[91] 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

[92] 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

[93] G. HALMANS, K. POHL.*Communicating the Variability of a Software-Product Family to Customers*, in "Software and System Modeling", 2003, vol. 2, n^o 1, p. 15-36

[94] C. HARDEBOLLE, F. BOULANGER.*ModHel'X: A component-oriented approach to multi-formalism modeling*, in "Models in Software Engineering", Springer, 2008, p. 247–258

[95] M. HARMAN, B. F. JONES.*Search-based software engineering*, in "Information and Software Technology", 2001, vol. 43, n^o 14, p. 833–839

[96] 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

[97] 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

[98] J.-M. Jézéquel.*Model Driven Design and Aspect Weaving*, in "Journal of Software and Systems Modeling (SoSyM)", may 2008, vol. 7, n^{o} 2, p. 209–218, http://www.irisa.fr/triskell/publis/2008/Jezequel08a.pdf

[99] 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

[100] J. Kramer, J. Magee.*Self-Managed Systems: an Architectural Challenge*, in "Future of Software Engineering", IEEE, 2007, p. 259–268

[101] K.-K. Lau, P. V. Elizondo, Z. Wang.*Exogenous connectors for software components*, in "Component-Based Software Engineering", Springer, 2005, p. 90–106

[102] P. McMinn.*Search-based software test data generation: a survey*, in "Software Testing, Verification and Reliability", 2004, vol. 14, n^{o} 2, p. 105–156

[103] J. Meekel, T. B. Horton, C. Mellone.*Architecting for Domain Variability*, in "ESPRIT ARES Workshop", 1998, p. 205-213

[104] A. M. Memon.*An event-flow model of GUI-based applications for testing*, in "Software Testing, Verification and Reliability", 2007, vol. 17, n^{o} 3, p. 137–157

[105] 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

[106] 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

[107] 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

[108] 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

[109] 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

[110] L. M. Northrop.*SEI's Software Product Line Tenets*, in "IEEE Softw.", 2002, vol. 19, n^{o} 4, p. 32–40

[111] 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

[112] 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^{o} 2, p. 128–145

[113] D. L. PARNAS.*On the Design and Development of Program Families*, in "IEEE Trans. Softw. Eng.", 1976, vol. 2, n^{o} 1, p. 1–9

[114] 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, n^{o} 4, p. 252–268

[115] 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

[116] B. RANDELL.*System structure for software fault tolerance*, in "Software Engineering, IEEE Transactions on", 1975, n^{o} 2, p. 220–232

[117] M. RINARD.*Obtaining and reasoning about good enough software*, in "Proceedings of Annual Design Automation Conference (DAC)", 2012, p. 930-935

[118] 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

[119] P. RUNESON, M. HÖST.*Guidelines for conducting and reporting case study research in software engineering*, in "Empirical Software Engineering", 2009, vol. 14, n^{o} 2, p. 131–164

[120] D. SCHMIDT.*Guest Editor's Introduction: Model-Driven Engineering*, in "IEEE Computer", 2006, vol. 39, n^{o} 2, p. 25–31

[121] F. SHULL, J. SINGER, D. I. SJBERG.*Guide to advanced empirical software engineering*, Springer, 2008

[122] 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

[123] J. STEEL, J.-M. JÉZÉQUEL.*On Model Typing*, in "Journal of Software and Systems Modeling (SoSyM)", December 2007, vol. 6, n^{o} 4, p. 401–414, http://www.irisa.fr/triskell/publis/2007/Steel07a.pdf

[124] C. SZYPERSKI, D. GRUNTZ, S. MURER.*Component software: beyond object-oriented programming*, Addison-Wesley, 2002

[125] J.-C. TRIGAUX, P. HEYMANS.*Modelling variability requirements in Software Product Lines: a comparative survey*, FUNDP Namur, 2003

[126] M. UTTING, B. LEGEARD.*Practical model-based testing: a tools approach*, Morgan Kaufmann, 2010

[127] P. VROMANT, D. WEYNS, S. MALEK, J. ANDERSSON.*On interacting control loops in self-adaptive systems*, in "SEAMS 2011", ACM, 2011, p. 202–207

[128] 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^o 1, p. 20–34

[129] 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

[130] 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

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
A3.1.10. - Heterogeneous data
A3.2.3. - Inference
A3.2.4. - Semantic Web
A3.2.5. - Ontologies
A3.2.6. - Linked data
A3.3.3. - Big data 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
A9.8. - Reasoning

Other Research Topics and Application Domains:

B1.1.2. - Molecular and cellular biology
B1.1.7. - Bioinformatics
B1.1.10. - Systems and synthetic biology
B2.2.3. - Cancer
B2.2.5. - Immune system diseases

1. Team, Visitors, External Collaborators

Research Scientists

Samuel Blanquart [Inria, Researcher]
François Coste [Inria, Researcher]
Anne Siegel [CNRS, Senior Researcher, HDR]

Faculty Members

Olivier Dameron [Team leader, Univ de Rennes I, Associate Professor, HDR]
Emmanuelle Becker [Univ de Rennes I, Associate Professor]
Catherine Belleannée [Univ de Rennes I, Associate Professor]

External Collaborators

François Moreews [INRA]
Denis Tagu [INRA, HDR]
Nathalie Théret [INSERM, HDR]

Technical Staff

Méziane Aite [CNRS, Engineer]
Arnaud Belcour [CNRS, Engineer, until Jul 2019]
Xavier Garnier [Inria, Engineer]
Jeanne Got [CNRS, Engineer]

PhD Students

Arnaud Belcour [Inria, PhD Student, from Oct 2019]
Lucas Bourneuf [Univ de Rennes I, PhD Student, until Nov 2019]
Mael Conan [Univ de Rennes I, PhD Student, UMR IRSET Lab]
Olivier Dennler [INSERM, PhD Student, from Oct 2019]
Nicolas Guillaudeux [Univ de Rennes I, PhD Student]
Virgilio Kmetzsch Rosa E Silva [Inria, PhD Student, from Oct 2019]
Marine Louarn [Univ de Rennes I, PhD Student]
Hugo Talibart [Univ de Rennes I, PhD Student]
Pierre Vignet [Univ de Rennes I, PhD Student, UMR IRSET Lab]
Meline Wéry [Univ de Rennes I, PhD Student, SANOFI R&D]

Post-Doctoral Fellows

Celia Biane-Fourati [Inria, Post-Doctoral Fellow]
Clémence Frioux [CNRS, Post-Doctoral Fellow, until May 2019]
Vijay Ingalalli [Univ de Rennes I, Post-Doctoral Fellow, until Jan 2019]

Visiting Scientist

Oumarou Abdou Arbi [Université de Maradi, Niger, until Feb 2019]

Administrative Assistant

Marie Le Roic [Inria, Administrative Assistant]

2. Overall Objectives

2.1. Overall objectives

Bioinformatics context: from life data science to functional information about biological systems and unconventional species. Sequence analysis and systems biology both consist in the interpretation of biological information at the molecular level, that concern mainly intra-cellular compounds. Analyzing genome-level information is the main issue of **sequence analysis**. The ultimate goal here is to build a full catalogue of bio-products together with their functions, and to provide efficient methods to characterize such bio-products in genomic sequences. In regards, contextual physiological information includes all cell events that can be observed when a perturbation is performed over a living system. Analyzing contextual physiological information is the main issue of **systems biology**.

For a long time, computational methods developed within sequence analysis and dynamical modeling had few interplay. However, the emergence and the democratization of new sequencing technologies (NGS, metagenomics) provides information to link systems with genomics sequences. In this research area, the Dyliss team focuses on linking genomic sequence analysis and systems biology. **Our main applicative goal in biology is to characterize groups of genetic actors that control the phenotypic response of species when challenged by their environment**. **Our main computational goals are to develop methods for analyzing the dynamical response of a biological system, modeling and classifying families of gene products with sensitive and expressive languages, and identifying the main actors of a biological system within static interaction maps**. We first 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 perturbations, sequencing...). We then rely on symbolic methods (semantic web technologies, solving combinatorial optimization problems, formal classification) to compute the main features of the space of admissible models.

Computational challenges. The main challenges we face are **data incompleteness and heterogeneity, leading to non-identifiability**. Indeed, we have observed that the biological systems that we consider cannot be uniquely identifiable. Indded, "omics" technologies have allowed the number of measured compounds in a systems to increase tremendously. However, it appears that the theoretical number of different experimental measurements required to integrate these compounds in a single discriminative model has increased exponentially with respect to the number of measured compounds. Therefore, according to the current state of knowledge, there is no possibility to explain the data with a single model. Our rationale is that biological systems will still remain non-identifiable for a very long time. In this context, we favor **the construction and the study of a space of feasible models or hypotheses** including known constraints and facts on a living system rather than searching for a single discriminative optimized model. We develop methods allowing a precise and exhaustive investigation of this space of hypotheses. With this strategy, we are in position of developing experimental strategies to progressively shrink the space of hypotheses and gain in the understanding of the system.

Bioinformatics challenges. Our objectives in computer sciences are developed within the team in order to fit with three main bioinformatics challenges (1) data-science and knowledge-science for life sciences (see Sec. 3.2) (2) Understanding metabolism (see Sec. 3.3) (3) Characterizing regulatory and signaling phenotypes (see Sec. 3.4).

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. They 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].

3. Research Program

3.1. Computer science – symbolic artificial intelligence

We develop methods that use an explicit representation of the relationships between heterogeneous data and knowledge in order to construct a space of hypotheses. Therefore, our objectives in computer science is mainly to develop accurate representations (oriented graphs, Boolean networks, automata, or expressive grammars) to iteratively capture the complexity of a biological system.

Integrating data with querying languages: Semantic web for life sciences The first level of complexity in the data integration process consists in confronting heterogeneous datasets. Both the size and the heretogeneity of life science data make their integration and analysis by domain experts impractical and prone to the streetlight effect (they will pick up the models that best match what they know or what they would like to discover). Our first objective involves the formalization and management of knowledge, that is, the explicitation of relations occurring in structured data. In this setting, our main goal is to facilitate and optimize the integration of Semantic Web resources with local users data by relying on the implicit data scheme contained in biological data and Semantic Web resources.

Reasoning over structured data with constraint-based logical paradigms Another level of complexity in life science integration is that very few paradigms exist to model the behavior of a complex biological system. This leads biologists to perform and formulate hypotheses in order to interpret their data. Our strategy is to interpret such hypotheses as combinatorial optimization problems allowing to reduce the family of models compatible with data. To that goal, we collaborate with Potsdam University in order to use and challenge the most recent developments of Answer Set Programming (ASP) [58], a logical paradigm for solving constraint satisfiability and combinatorial optimization issues. Our goal is therefore to provide scalable and expressive formal models of queries on biological networks with the focus of integrating dynamical information as explicit logical constraints in the modeling process.

Characterizing biological sequences with formal syntactic models Our last goal is to identify and characterize the function of expressed genes in non-model species, such as enzymes and isoforms functions in biological networks or specific functional features of metagenomic samples. These are insufficiently precise because of the divergence of biological sequences, the complexity of molecular structures and biological processes, and the weak signals characterizing these elements. Our goal is therefore to develop accurate formal syntactic models (automata, grammars, abstract gene models) enabling us to represent sequence conservation, sets of short and degenerated patterns and crossing or distant dependencies. This requires both to determine classes of formal syntactic models allowing to handle biological complexity, and to automatically characterize the functional potential embodied in biological sequences with these models.

3.2. Scalable methods to query data heterogenity

Confronted to large and complex data sets (raw data are associated with graphs depicting explicit or implicit links and correlations) almost all scientific fields have been impacted by the *big data issue*, especially genomics and astronomy [67]. In our opinion, life sciences cumulates several features that are very specific and prevent the direct application of big data strategies that proved successful in other domains such as experimental physics: the existence of **several scales of granularity** from microscopic to macroscopic and the associated issue of dependency propagation, datasets **incompleteness and uncertainty** including highly **heterogeneous** responses to a perturbation from one sample to another, and highly fragmented sources of information that **lacks interoperability** [57]. To explore this research field, we use techniques from symbolic data mining (Semantic Web technologies, symbolic clustering, constraint satisfaction and grammatical modelling) to take into account those life science features in the analysis of biological data.

3.2.1. Research topics

Facilitating data integration and querying The quantity and inner complexity of life science data require semantically-rich analysis methods. A major challenge is then to combine data (from local project as well as from reference databases) and symbolic knowledge seamlessly. Semantic Web technologies (RDF for annotating data, OWL for representing symbolic knowledge, and SPARQL for querying) provide a relevant framework, as demonstrated by the success of Linked (Open) Data [44]. However, life science end users (1) find it difficult to learn the languages for representing and querying Semantic Web data, and consequently (2) miss the possibility they had to interact with their tabulated data (even when doing so was exceedingly slow and tedious). Our first objective in this axis is to develop accurate abstractions of datasets or knowledge repositories to facilitate their exploration with RDF-based technologies.

Scalability of semantic web queries. A bottleneck in data querying is given by the performance of federated SPARQL queries, which must be improved by several orders of magnitude to allow current massive data to be analyzed. In this direction, our research program focuses on the combination of *linked data fragments* [68], query properties and dataset structure for decomposing federated SPARQL queries.

Building and compressing static maps of interacting compounds A final approach to handle heterogeneity is to gather multi-scale data knowledge into functional static map of biological models that can be analyzed and/or compressed. This requires to linking genomics, metabolomics, expression data and protein measurement of several phenotypes into unified frameworks. In this direction, our main goal is to develop families of constraints, inspired by symbolic dynamical systems, to link datasets together. We currently focus on health (personalized medicine) and environmental (role of non-coding regulations, graph compression) datasets.

3.2.2. Associated software tools

AskOmics platform *AskOmics* is an integration and interrogation software for linked biological data based on semantic web technologies [url]. *AskOmics* aims at bridging the gap between end user data and the Linked (Open) Data cloud (LOD 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. It helps 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. An originality is the graphical interface that allows any dataset to be integrated in a local RDF datawarehouse and SPARQL query to be built transparently and iteratively by a non-expert user.

FinGoc-tools *The FinGoc tools* allow filtering interaction networks with graph-based optimization criteria in order to elucidate the main regulators of an observed phenotype. The main added-value of these tools is to make explicit the criteria used to highlight the role of the main regulators.(1) The KeyRegulatorFinder package searches key regulators of lists of molecules (like metabolites, enzymes or genes) by taking advantage of knowledge databases in cell metabolism and signaling [package]. (2) The PowerGrasp python package implements graph compression methods oriented toward visualization, and based on power graph analysis [package]. (3) The iggy package enables the repairing of an interaction graph with respect to expression data. [Python package]

3.3. Metabolism: from enzyme sequences to systems ecology

Our researches in bioinformatics in relation with metabolic processes are driven by the understanding of non-model (eukaryote) species. Their metabolism have acquired specific features that we wish to identify with computational methods. To that goal, we combine sequence analysis with metabolic network analysis, with the final goal to understand better the metabolism of communities of organisms.

3.3.1. Research topics

Genomic level: characterizing enzymatic functions of protein sequences Precise characterization of functional proteins, such as enzymes or transporters, is a key to better understand and predict the actors involved in a metabolic process. In order to improve the precision of functional annotations, we develop machine learning approaches taking a sample of functional sequences as input to infer a grammar representing their key syntactical characteristics, including dependencies between residues. Our first goal is to enable an automatic semi-supervised refinement of enzymes classification [6] by combining the Protomata-Learner [50] framework - which captures local dependencies - with formal concept analysis. More challenging, we are exploring the learn of grammars representing long-distance dependencies such as those exhibited by contacts of amino-acids that are far in the sequence but close in the 3D protein folding.

System level: enriching and comparing metabolic networks for non-model organisms Non-model organisms are associated with often incomplete and poorly annotated sequences, leading to draft networks of their metabolism which largely suffer from incompleteness. In former studies, the team has developed several methods to improve the quality of eukaryotes metabolic networks, by solving several variants of the so-called *Metabolic Network gap-filling problem* with logical programming approaches [10], [9]. The main drawback of these approaches is that they cannot scale to the reconstruction and comparison of families of metabolic networks. Our main objective is therefore to develop new tools for the comparison of species strains at the metabolic level.

Consortium level: exploring the diversity of community consortia A new emerging field is system ecology, which aims at building predictive models of species interactions within an ecosystem for deciphering cooperative and competitive relationships between species [56]. This field raises two new issues (1) uncertainty on the species present in the ecosystem and (2) uncertainty about the global objective governing an ecosystem. To address these challenges, our first research focus is the inference of metabolic exchanges and relationships for transporter identification, based on our expertise in metabolic network gap-filling. A second very challenging focus is the prediction of transporters families by obtaining refined characterization of transporters, which are quite unexplored apart from specific databases [65].

3.3.2. Associated software tools

Protomata[url] is a machine learning suite for the inference of automata characterizing (functional) families of proteins at the sequence level. It provides programs to build a new kind of sequences alignments (said partial and local), learn automata and search for new family members in sequence databases. By enabling to model dependencies between positions, automata are more expressive than classical tools (PSSMs, Profile HMMs, or Prosite Patterns) and are well suited to predict new family members with a high specificity. This suite is for instance embedded in the cyanolase database [50] to automate its updade and was used for refining the classification of HAD enzymes [6].

AuReMe workspace is designed for tractable reconstruction of metabolic networks [url]. The toolbox allows for the Automatic Reconstruction of Metabolic networks based on the combination of multiple heterogeneous data and knowledge sources [1]. The main added-values are the inclusion of graph-based tools relevant for the study of non-classical organisms (Meneco and Menetools packages), the possibility to trace the reconstruction and curation procedures (Padmet and Padmet-utils packages), and the exploration of reconstructed metabolic networks with wikis (wiki-export package, see: [url]. It also generated outputs to explore resulting networks with Askomics. It has been used for reconstructing metabolic networks of micro and macro-algae [62], extremophile bacteria [52] and communities of organisms [4].

Mpwt is a Python package for running Pathway Tools [url] on multiple genomes using multiprocessing. Pathway Tools is a comprehensive systems biology software system that is associated with the BioCyc database collection [url]. Pathway Tools is very used for reconstructing metabolic networks.

Metage2metabo is a Python tool to perform graph-based metabolic analysis starting from annotated genomes (reference genomes or metagenome-assembled genomes). It uses Mpwt to reconstruct metabolic networks for a large number of genomes. The obtained metabolic networks are then analyzed individually and collectively in order to get the added value of metabolic cooperation in microbiota over individual metabolism and to identify and screen interesting organisms among all.

3.4. Regulation and signaling: detecting complex and discriminant signatures of phenotypes

On the contrary to metabolic networks, regulatory and signaling processes in biological systems involves agents interacting at different granularity levels (from genes, non-coding RNAs to protein complexes) and different time-scales. Our focus is on the reconstruction of large-scale networks involving multiple scales processes, from which controllers can be extracted with symbolic dynamical systems methods. A particular attention is paid to the characterization of products of genes (such as isoform) and of perturbations to identify discriminant signature of pathologies.

3.4.1. Research topics

Genomic level: characterizing gene structure with grammatical languages and conservation information The subject here is to accurately represent gene structure, including intron/exon structure, for predicting the products of genes, such as isoform transcripts, and comparing the expression potential of a eukaryotic gene according to its context (e.g. tissue) or according to the species. Our approach consists in designing grammatical and comparative-genomics based models for gene structures able to detect heterogeneous functional sites (splicing sites, regulatory binding sites...), functional regions (exons, promotors...) and global constraints (translation into proteins) [46]. Accurate gene models are defined by identifying general constraints shaping gene families and their structures conserved over evolution. Syntactic elements controlling gene expression (transcription factor binding sites controlling transcription; enhancers and silencers controlling splicing events...), i.e. short, degenerated and overlapping functional sequences, are modeled by relying on the high capability of SVG grammars to deal with structure and ambiguity [66].

System level: extracting causal signatures of complex phenotypes with systems biology frameworks The main challenge we address is to set up a generic formalism to model inter-layer interactions in large-scale biological networks. To that goal, we have developed several types of abstractions: multi-experiments framework to learn and control signaling networks [11], multi-layer reactions in interaction graphs [47], and multi-layer information in large-scale Petri nets [43]. Our main issues are to scale these approaches to standardized large-scale repositories by relying on the interoperable Linked Open Data (LOD) resources and to enrich them with ad-hoc regulations extracted from sequence-based analysis. This will allow us to characterize changes in system attractors induced by mutations and how they may be included in pathology signatures.

3.4.2. Associated software tools

Logol software is designed for complex pattern modelling and matching [url]. It is a swiss-army-knife for pattern matching on DNA/RNA/Protein sequences, based on expressive patterns which consist in a complex combination of motifs (such as degenerated strings) and structures (such as imperfect stem-loop ou repeats) [2]. *Logol* key features are the possibilities (i) to divide a pattern description into several sub-patterns, (ii) to model long range dependencies, and (iii) to enable the use of ambiguous models or to permit the inclusion of negative conditions in a pattern definition. Therefore, *Logol* encompasses most of the features of specialized tools (Vmatch, Patmatch, Cutadapt, HMM) and enables interplays between several classes of patterns (motifs and structures), including stem-loop identification in CRISPR.

Caspo software Cell ASP Optimizer (*Caspo*) constitutes a pipeline for automated reasoning on logical signaling networks (learning, classifying, designing experimental perturbations, identifying controllers, take time-series into account) [url]. The software handles inherent experimental noise by enumerating all different logical networks which are compatible with a set of experimental observations [11]. The main advantage is that it enables a complete study of logical network without requiring any linear constraint programs.

Cadbiom package aims at building and analyzing the asynchronous dynamics of enriched logical networks [url] It is based on Guarded transition semantic and allows synchronization events to be investigated in large-scale biological networks [43]. For instance, it was designed to allow controler of phenotypes in large-scale knowledge databases (PID) to be curated and analyzed [5].

4. Application Domains

4.1. Application fields in biology

In terms of transfer and societal impact, we consider that our role is to develop fruitful collaborations with laboratories of biology in order to consolidate their studies by a smart use of our tools and prototypes and to generate new biological hypotheses to be tested experimentally.

Marine Biology: seaweed enzymes and metabolism & sea-urchin cell-cycle. Our main field of study is **marine biology**, as it is a transversal field covering challenges in integrative biology, dynamical systems and sequence analysis. 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 seaweed metabolism *E. Siliculosus* to be deciphered [62], [53]. The study of the *HAD* superfamily of proteins thanks to partial local alignments, produced by *Protomata* tools, allows sub-families to be deciphered and classified, and the metabolic map reconstructed with *Meneco* enabled the reannotation of 56 genes within the *E. siliculosus* genome. These approaches also shed light on evolution of metabolic processes. As a further study, we reconstructed the metabolic network of a symbiot bacterium *Ca. P. ectocarpi* [55] and used this reconstructed network to decipher interactions within the algal-bacteria holobiont, revealing several candidates metabolic pathways for algal-bacterial interactions. Similarily, our analyses suggest that the bacterium *Ca. P. ectocarpi* is able to provide both β-alanine and vitamin B5 to the seaweed via the phosphopantothenate biosynthesis pathway [63].

Micro-biology: elucidating the functioning of extremophile consortiums of bacteria. In this application field, our main issue is the understanding of bacteria living in extreme environments, mainly in collaboration with the group of bioinformatics at Universidad de Chile (co-funded by the Center of Mathematical Modeling, the Center of Regulation Genomics and Inria-Chile). In order to elucidate the main characteristics of these bacteria, our integrative methods were developed to identify the main groups of regulators for their specific response in their living environment. The integrative biology tools *Meneco*, *Lombarde* and *Shogen* have been designed in this context. In particular, genome-scale metabolic network been recently reconstructed and studied with the *Meneco* and *Shogen* approaches, especially on bacteria involved in biomining processes [48] and in Salmon pathogenicity [52].

Agriculture and environmental sciences: upstream controllers of cow, pork and pea-aphid metabolism and regulation. In this application field, our goal is to propose methods to identify regulators of very complex phenotypes related to environmental issues. Our work on the identification of upstream regulators within large-scale knowledge databases (prototype *KeyRegulatorFinder*) [47] and on semantic-based analysis of metabolic networks [45] was very valuable for interpreting differences of gene expression in pork meat [60] and figure out the main gene-regulators of the response of porks to several diets [59]. 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 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.

Health: deciphering pathways involved in the TGF-β signalling network. TGF-β is a multifunctional cytokine that regulates mammalian development, differentiation, and homeostasis with both benefical anti-tumor effect [49] and pro-tumor effect [61]. Deciphering protumor versus antitumor signaling requires to take into account a system-wide view and develop predictive models for therapeutic benefit. For that purpose we developed *Cadbiom* and identified gene networks associated with innate immune response to viral infection that combine TGF-β and interleukine signaling pathways [43], [51].

5. Highlights of the Year

5.1. Highlights of the Year

The AuReMe software for metabolic network reconstruction has been selected for the Service Delivery Plan of the French Institute of Bioinformatics (IFB).

5.1.1. Awards

Lucas Bourneuf is the World champion of man vs. machine challenge of the Angry Birds AI competition (during IJCAI) where humans can challenge the four best AI agents. Note that Lucas was the human, not the AI and that there is no direct connection with his PhD project. World champion nonetheless!

Nicolas Guillaudeux (with Grégoire Siekaniec from the GenScale team) won the public's prize at the short scientific film festival "Sciences en cour[t]s" for their movie about Nicolas's PhD thesis.

6. New Software and Platforms

6.1. AskOmics

Convert tabulated data into RDF and create SPARQL queries intuitively and "on the fly".

KEYWORDS: RDF - SPARQL - Querying - Graph - LOD - Linked open data

FUNCTIONAL 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) to be loaded in a RDF triplestore and then be transparently and interactively queried. AskOmics is made of three software blocks: (1) a web interface for data import, allowing the creation of a local triplestore from user's datasheets and standard data, (2) an interactive web interface allowing "à la carte" query-building, (3) a server performing interactions with local and distant triplestores (queries execution, management of users parameters).

NEWS OF THE YEAR: (1) migration to github, (2) complete re-engineering for cleaning the successive layers and accomodating further extensions, (3) integration of the Corese triplestore (https://corese.inria.fr/) in addition to fuseki and virtuoso, (4) improved user interface, (5) capability to save queries for sharing and reusing them, (6) automatic generation of askomics-compliant graph of entity types (abstraction), (7) capability to use askomics to query remote endpoints (including Uniprot and neXtProt), (8) support for federated queries involving remote endpoints and local data

- Authors: Charles Bettembourg, Xavier Garnier, Anthony Bretaudeau, Fabrice Legeai, Olivier Dameron, Olivier Filangi and Yvanne Chaussin
- Partners: Université de Rennes 1 - CNRS - INRA
- Contact: Olivier Dameron
- URL: https://github.com/askomics/askomics

6.2. AuReMe

Automatic Reconstruction of Metabolic networks

KEYWORDS: Workflow - Bioinformatics - Metabolic networks - Omic data - Toolbox - Data management

FUNCTIONAL DESCRIPTION: AuReMe enables the reconstruction of metabolic networks from different sources based on sequence annotation, orthology, gap-filling and manual curation. The metabolic network is exported as a local wiki allowing to trace back all the steps and sources of the reconstruction. It is highly relevant for the study of non-model organisms, or the comparison of metabolic networks for different strains or a single organism.

Five modules are composing AuReMe: 1) The Model-management PADmet module allows manipulating and traceing all metabolic data via a local database. 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 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. 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.

RELEASE FUNCTIONAL DESCRIPTION: - Reworking padmet and padmet-utils to allow full-python workflow in the future - Adding new script padmet-utils/exploration/prot2genome with exonerate - Fixing minor errors

NEWS OF THE YEAR: (1) Pantograph replaced by OrthoFinder (2) Create a readthedocs for AuReMe, padmet and padmet-utils (3) Reworking padmet and padmet-utils to allow full python workflow (4) Adding new script padmet-utils/exploration/prot2genome with exonerate (5) Modify template data structure (6) Fixing errors

- Participants: Marie Chevallier, Meziane Aite, Guillaume Collet, Nicolas Loira, Sylvain Prigent, Jeanne Cambefort, Anne Siegel and Alejandro Maass
- Partner: University of Chile
- Contact: Meziane Aite
- Publication: Traceability, reproducibility and wiki-exploration for "à-la-carte" reconstructions of genome-scale metabolic models
- URL: http://aureme.genouest.org/

6.3. biseau

KEYWORDS: ASP - Answer Set Programming - Graph - Formal concept analysis

SCIENTIFIC DESCRIPTION: Use ASP as a Domain Specific Language to specify dot-based visualizations.

NEWS OF THE YEAR: First release.

- Contact: Lucas Bourneuf
- Publication: An Answer Set Programming Environment for High-Level Specification and Visualization of FCA
- URL: https://gitlab.inria.fr/lbourneu/biseau

6.4. Metage2Metabo

KEYWORDS: Metabolic networks - Microbiota - Metagenomics - Workflow

FUNCTIONAL DESCRIPTION: Metabolic networks are graphs which nodes are compounds and edges are biochemical reactions. To study the metabolic capabilities of microbiota, Metage2Metabo uses multiprocessing to reconstruct metabolic networks at large-scale. The individual and collective metabolic capabilities (number of compounds producible) are computed and compared. From these comparisons, a set of compounds only producible by the community is created. These newly producible compounds are used to find minimal communities that can produce them. From these communities, the keytstone species in the production of these compounds are identified.

NEWS OF THE YEAR: First release.

- Contact: Anne Siegel
- Publication: Metage2Metabo: metabolic complementarity applied to genomes of large-scale microbiotas for the identification of keystone species
- URL: https://github.com/AuReMe/metage2metabo

6.5. Pathmodel

KEYWORDS: ASP - Answer Set Programming - Metabolic networks - Metabolic Pathway Drift - Bioinformatics - Systems Biology - Metabolomics

SCIENTIFIC DESCRIPTION: This tool is a prototype of the Metabolic Pathway Drift concept. This concept states that metabolic pathways undergo substantial turnover. The reactions involved in a pathway can change between species (change in reaction order or replacement of an enzyme by another one). Another goal of this tool is to linked genomics and metabolomics data. To implement this concept, Pathmodel uses the Answer Set Programming language. The input are the reactants and products involved in the pathway, known reactions occurring between these molecules, known m/z ratio, known domains shard by these molecules, an initial molecule and a goal molecule. Using these data, Pathmodel will infer reactions between molecules to reach the goal molecule using the known reactions. The result consist of potential alternative pathways for the studied organism.

FUNCTIONAL DESCRIPTION: A metabolic pathway is a series of biochemical reactions. These reactions modify metabolites in order to synthesize a new metabolite or to produce energy. One difficulties when dealing with pathways in non-model organism is their incomplete conservation during evolution. To deal with this problem, we developed a prototype inferring new biochemical reactions using reactions and metabolites from known metabolic pathways and metabolomics data. This method produce alternative pathway that could occurred in the species of interest.

RELEASE FUNCTIONAL DESCRIPTION: Fix an issue with test data.

NEWS OF THE YEAR: (1) Add a container in Singularity Hub (https://singularity-hub.org/collections/3758). (2) Rewrite data files (sterol and MAA). (3) Add creation of pictures of new molecules from MZ. (4) Add new output files to ease understanding of PathModel output. (5) Rewrite the Readme.

- Participants: Arnaud Belcour, Jacques Nicolas, Gabriel Markov and Anne Siegel
- Partner: Station Biologique de Roscoff
- Contact: Anne Siegel
- Publication: Inferring biochemical reactions and metabolite structures to cope with metabolic pathway drift
- URL: https://github.com/pathmodel

6.6. CADBIOM

Computer Aided Design of Biological Models

KEYWORDS: Health - Biology - Biotechnology - Bioinformatics - Systems Biology

FUNCTIONAL DESCRIPTION: The Cadbiom software provides a formal framework to help the modeling of biological systems such as cell signaling network with Guarder Transition Semantics. It allows synchronization events to be investigated in biological networks among large-scale network in order to extract signature of controllers of a phenotype. 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, since the translations of the large-scale PID repository (about 10,000 curated interactions) have been translated into the Cadbiom formalism.

NEWS OF THE YEAR: - Comprehensive command line to run the calculations and analyze the generated results. - Module designed to produce models through the interpretation of various databases or ontologies, formalized according to the BioPAX standard. - Update of the site and the documentation.

We recently developed a framework that integrates an updated version of the CADBIOM core software and visualization tools. We provided a command line interface allowing users to translate the interactions between biomolecules described in data sources in BioPAX format into the formalism based on the guarded transitions used by CADBIOM. The command line also makes it easy to search for scenarios based on the constraints of a model, to compare scenarios with each other and to visualize interaction graphs facilitating the biologist's expertise (Vignet et al 2019 JOBIM)

- Participants: Geoffroy Andrieux, Michel Le Borgne, Nathalie Theret, Nolwenn Le Meur, Pierre Vignet and Anne Siegel
- Contact: Anne Siegel
- URL: http://cadbiom.genouest.org

7. New Results

7.1. Scalable methods to query data heterogenity

Participants: Emmanuelle Becker, Lucas Bourneuf, Olivier Dameron, Xavier Garnier, Vijay Ingalalli, Marine Louarn, Yann Rivault, Anne Siegel.

Increasing life science resources re-usability using Semantic Web technologies [*E. Becker, O. Dameron, X. Garnier, V. Ingallali, M. Louarn, Y. Rivault, A. Siegel*] [25], [18], [29], [31], [23], [27], [28]. Our work was focused on assessing to what extent Semantic Web technologies also facilitate reproducibility and reuse of life sciences studies involving pipelines that compute associations between entities according to intermediary relations and dependencies.

- We followed on 2018 action exploratoire Inria by studying possible optimizations for federated SPARQL queries [31]
- We considered a case-study in systems biology ([Regulatorycircuits link]), which provides tissue-specific regulatory interaction networks to elucidate perturbations across complex diseases. We relied on this structure and used Semantic Web technologies (i) to integrate the Regulatory Circuits data, and (ii) to formalize the analysis pipeline as SPARQL queries. Our result was a 335,429,988 triples dataset on which two SPARQL queries were sufficient to extract each single tissue-specific regulatory network.

- A second case-study concerned public health data for reusing electronic health data, selecting patients, identifying specific events and interpreting results typically requires biomedical knowledge [64]. We developed the queryMed R package [18], [29]. It aims to facilitate the integration of medical and pharmacological knowledge stored in formats compliant with the Linked Data paradigm (e.g. OWL ontologies and RDF datasets) into the R statistical programming environment. We showed that linking a medical database of 1003 critical limb ischemia (CLI) patients to ontologies allowed us to identify all the drugs prescribed for CLI and also to detect one contraindicated prescription for one patient. We also investigated temporal models of care sequences for the exploration of medico-administrative data as part of Johanne Bakalara's PhD, supervized with Thomas Guyet (Lacodam) and Emmanuel Oger (Repères).
- We pursued the development of AskOmics [27]. Version 3 adds the capability to generate the graph of entity types (aka abstraction) from typed RDF datasets, improved management of entity hierarchies and support for federated queries on external SPARQL endpoints such as UniProt and neXtProt.

Graph compression and analysis [*L. Bourneuf*] [26], [24]. Because of the increasing size and complexity of available graph structures in experimental sciences like molecular biology, techniques of graph visualization tend to reach their limit.

- We developed the Biseau approach, a programming environment aiming at simplifying the visualization task. Biseau takes advantage of Answer Set Programming and shows as a use-case how Formal Concept Analysis can be efficiently described at the level of its properties, without needing a costly development process. It reproduces the core results of existing tools like LatViz or In-Close.
- We formalized a graph compression search space in order to provide approximate solutions to the NP-complete problem of computing a lossless compression of the graph based on the search of cliques and bicliques. Our conclusion is that the search for graph compression can be usefully associated with the search for patterns in a concept lattice and that, conversely, confusing sets of objects and attributes brings new interesting problems for FCA.

7.2. Metabolism: from enzyme sequences to systems ecology

Participants: Méziane Aite, Arnaud Belcour, Mael Conan, François Coste, Clémence Frioux, Jeanne Got, Anne Siegel, Hugo Talibart.

Modelling proteins with long distance dependencies [*F. Coste, H. Talibart*] [15], [30], [30]

- We proposed to use information on protein contacts to train probabilistic context-free grammars representing families of protein sequences. We developed the theory behind the introduction of contact constraints in maximum-likelihood and contrastive estimation schemes and implemented it in a machine learning framework for protein grammars. Evaluation showed high fidelity of grammatical descriptors to protein structures, improved precision in recognizing sequences and the ability to model a meta-family of proteins that could not be modeled by classical approaches [15].
- We then investigated the problem of modeling proteins with crossing dependencies. Motivated by their success on contact prediction, we propose to use Potts models for the purposes of modeling proteins and searching. We developed ComPotts a tool for optimal alignment and comparison of Potts models, enabling to take into account the coevolution of residues for the search of protein homologs [30], [40].

Large-scale eukaryotic metabolic network reconstruction [*A. Siegel, C. Frioux, M. Aite, A. Belcour, J. Got, N. Théret, M. Conan*] [17], [14], [38]. Metabolic network reconstruction has attained high standards but is still challenging for complex organisms such as eukaryotes.

- *Large-scale eukaryotic metabolic network reconstruction*: We participated to the reconstruction of a genome–scale metabolic network for the brown Algae *Saccharina japonica* and *Cladosiphon okamuranus* in order to shed light of the specificities on the carotenoid biosynthesis Pathway.

- *Metabolic pathway inference from non genomic data*: We designed methods for the identification of metabolic pathways for which enzyme information is not precise enough. As an application study, we focused on Heterocyclic Aromatic Amines (HAAs), which are environmental and food contaminants classified as probable carcinogens. Our approach based on a refinement of molecular predictions with enzyme activity scores allows to accurately predict HAAs biotransformation and their potentials DNA reactive compounds [54].

Systems ecology: design of microbial consortia [*C. Frioux, A. Belcour, J. Got, M. Aite, A. Siegel*] [21], [22], [34], [33].

- We participated to the application of our methods to algal-microbial consortia, with good preliminary results, and presented them as an invited conference [22].

7.3. Regulation and signaling: detecting complex and discriminant signatures of phenotypes

Participants: Catherine Belleannée, Célia Biane-Fourati, Samuel Blanquart, Olivier Dameron, Maxime Folschette, Nicolas Guillaudeux, Marine Louarn, François Moreews, Anne Siegel, Nathalie Théret, Pierre Vignet, Méline Wéry.

Creation of predictive functional signaling networks [*M. Folschette, N. Théret*] [16].

- Integrating genome-wide gene expression patient profiles with regulatory knowledge is a challenging task because of the inherent heterogeneity, noise and incompleteness of biological data. We proposed an automatic pipeline to extract automatically regulatory knowledge from pathway databases and generate novel computational predictions related to the state of expression or activity of biological molecules. We applied it in the context of hepatocellular carcinoma (HCC) progression, and evaluated the precision and the stability of these computational predictions. Our computational model predicted the shifts of expression of 146 initially non-observed biological components. Our predictions were validated at 88% using a larger experimental dataset and cross-validation techniques.

Experimental evidences of transcript predictions [*C. Belleannée, S. Blanquart, N. Guillaudeux*] [13].

- We designed comparative-genomics based models of gene structures through genes comparisons across species. These models enable to predict putative transcript isoforms in a species given the knowledge available in other species [46]. We recently published a first experimental validation of such a predicted transcriptome [13]. In this work, transcript isoforms of the human TRPM8 gene yield transcript predictions in the mouse TRPM8 gene, which are experimentally validated using targeted PCR in mouse tissues. This work also provides a first attempt to estimate origin of new isoforms during the gene evolution.
- In another collaboration with IGDR, we considered a multi-species gene comparison including human, mouse and dog [39]. This work reveals global trends of the gene isoform sets evolution, suggesting a extremely high plasticity of alternative transcription and alternative splicing propensities in those three species. This work moreover provides experimental evidences of the predicted transcripts based on public RNAseq data, highlighting the tissue specificity of isoform expression across species.

Formalizing and enriching phenotype signatures using Boolean networks [*C. Biane-Fourati, M. Wéry, A. Siegel, O. Dameron*] [20], [42], [22]

- We used Formal Concept Analysis as a symbolic bi-clustering techniques to classify and sort the steady states of a Boolean network according to biological signatures based on the hierarchy of the roles the network components play in the phenotypes. We applied our approach on a T helper lymphocyte (Th) differentiation network with a set of signatures corresponding to the sub-types of Th. This led to the identification and prediction of a new hybrid sub-type later confirmed by the literature.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. SANOFI: co-supervised PhD

Participant: Emmanuelle Becker.

This collaboration project is focused on the implementation of an integrative analysis framework based on semantic web technologies and reasoning in the framework of systemic lupus erythematosus pathology [42]. **CIFRE co-supervised Grant: Ph.D. funding. 2017-2020**

8.1.2. Theranexus: co-supervised internship

Participant: Pierre Beaudier.

This collaboration project was focused on assessing public databases' relevance for predicting potential drug combinations in central nervous system's pathologies [32]. It opened the perspective of a CIFRE PhD with Insiliance (under review by ANRT) **Theranexus funding. 2019**

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. MoDaL (Brittany and Pays de la Loire regions)

Participant: Olivier Dameron.

The MoDaL project is a federated project funded by BioGenOuest (Région Bretagne-Pays de la Loire) project involving scientists and engineers from IRISA/inria rennes (Genouest, Empenn, Dyliss) and the Institut du Thorax lab in Nantes. The project aims to decompartmentalize the resources dedicated to biomedical imaging and genetics. MoDaL focuses on i) establishing an inventory of the actors and infrastructures available at the inter-regional level, ii) proposing technological demonstrators that address the management, analysis and reuse of multi-infrastructure data (in-vivo, in-vitro and genomic imaging).

2019-2020. Total grant (hosted in the Empenn team): 100,000€.

9.1.2. PhenoMiR (European Maritime and Fisheries Fund)

Participant: Emmanuelle Becker.

The PhenoMiR project is a collaboration between Fishes Physiology and Genomics Laboratory (LPGP - INRAE), eight other laboratories of the INRAE, and the Dyliss team. Its objective is first (i) to settle the first complete repository of microRNAs for the trout, exploring different physiological and breeding conditions, and then (ii) to study the potentiality of some micro-RNAs to act as bio-markers of trout breeding and development condition. The Dyliss team is responsible for the design and development of the analysis pipeline of the genomics data, including the search of potential bio-markers.

2019-2022. Total grant: 495,000€. Dyliss grant: 33,000€.

9.1.3. UBIQUITIN

Participant: Emmanuelle Becker.

The Ubiquitin project is a collaboration between G. Rabut's team at the Institute of Developmental Biology of Rennes and the Dyliss team. It was funded as a cross-disciplinary emerging project by the University of Rennes 1.

G. Rabut's team is developing a new method to detect weak affinity protein-protein interactions based on protein complemantation with Luciferase. However the method may generate a very noisy signal depending on the *in-vivo* concentration of the partners. In the Ubiquitin project, we developped a R workflow to separate the signal from the noise in the experiments. As an application, this allowed us to decipher the intricate interplay between E2 and E3 enzymes during ubiquitination process in Yeast. This work was done during a master 2 internship. We are now continuing the project in two directions : (i) comparing the interactions identified with previously known databases, using web semantic technologies and ontologies describing protein interaction detection methods, and (ii) using formal classification to understand the structural properties of E2 and E3 that lead to their interactions. 2019-2020. Total grant: 7,000€. Dyliss grant: 4,200€.

9.1.4. *Ph.D. fundings from Université, Inria Rennes and Inserm*

The team benefits from PhD. theses fundings by Univ. Rennes (L. Bourneuf, 2016-2019 – H. Talibart, 2017-2020 – N. Guillaudeux, 2018-2021), by Inria (A. Belcour, 2019-2022 – V. Kmetzsch, 2019–2022), by Inserm (M. Louarn, 2017-2020, Inria-Inserm PhD Grant program), and by our collaborators from IRSET (M. Conan, 2017-2020 – P. Vignet, 2018-2020, O. Dennler, 2019–2022).

9.2. National Initiatives

9.2.1. *IDEALG (ANR/PIA-Biotechnology and Bioresource)*

Participant: Méziane Aite.

The project gathers 18 partners from Station Biologique de Roscoff (coordinator), CNRS, IFREMER, UEB, UBO, UBS, ENSCR, University of Nantes, INRA, AgroCampus, and the industrial field in order to foster biotechnology applications within the seaweed field. Dyliss is co-leader of the WP related to the establishment of a virtual platform for integrating omics studies on seaweed and the integrative analysis of seaweed metabolism. Major objectives are the building of brown algae metabolic maps, metabolic flux analysis and the selection of symbiotic bacteria for brown algae. We will also contribute to the prediction of specific enzymes (sulfatases and haloacid dehalogenase) [More details]. 2012–20. Total grant: 11M€. Dyliss grant: 534k€.

9.2.2. *TGFSysBio (ITMO Cancer)*

Participant: Olivier Dameron.

Partners are INSERM (coordinator) (IRSET, Univ. Rennes 1) CNRS (Dyliss team) and Inria (Antique, Paris). The TGFSYSBIO project aims at developing the first model of extracellular and intracellular TGF-beta system by combining a ruled-based modelling approach (kappa) and a Petri net modelling approach (cadbiom). 2015–18, extended in 2019. Total grant: 418k€. Dyliss grant: 129k€.

9.2.3. *Programs funded by Inria*

9.2.3.1. *IPL Neuromarkers*

Participant: Emmanuelle Becker.

This project involves mainly the Inria teams Aramis (coordinator) Dyliss, Genscale and Bonsai. The project aims at identifying the main markers of neurodegenerative pathologies through the production and the integration of imaging and bioinformatics data. Dyliss is in charge of facilitating the interoperability of imaging and bioinformatics data. In 2019 V. Kmetzsch started his PhD (supervized by E. Becker from Dyliss and O. Colliot from Aramis). 2017–20.

9.2.3.2. *Askomics (ADT)*

Participant: Olivier Dameron.

AskOmics [url] is a visual SPARQL query interface supporting both intuitive data integration and querying while avoiding the user to face most of the technical difficulties underlying RDF and SPARQL. The underlying motivation is that even though Linked (Open) Data now provide the infrastructure for accessing large corpora of data and knowledge, life science end-users seldom use them, nor contribute back their data to the LOD cloud by lack of technical expertise. AskOmics aims at bridging the gap between end users and the LOD cloud. 2018–2020.

9.3. European Initiatives

9.3.1. Collaborations in European Programs, Except FP7 & H2020

- Program: Polish National Science Center
- Project acronym: NCN 2016/21/B/ST6/02158
- Project title: Grammatical inference methods in classification of amyloidogenic proteins
- Duration: January 2017 - January 2020
- Coordinator: Olgierd Unold, Politechnika Wroclawska
- Other partners: Politechnika Wroclawska (Polland)
- Abstract: The objective is to develop the methods for induction of context-free and probabilistic grammars to describe a language matching amyloidogenic protein sequences.

9.3.2. Collaborations with Major European Organizations

Partner: Potsdam (Germany)

Title: Modeling combinatorial and hybrid optmization problems with Answer Set Programming

9.4. International Initiatives

9.4.1. Informal International Partners

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:** Oumarou Abdou-Arbi (University of Maradi)

9.5.1.1. Research Stays Abroad

- **Germany:** Maël Conan visited the Zentrum für Bioinformatik at Hamburg University with Prof. Johannes Kirchmair for 2 months. During this stay, he learned how to predict metabolism sites using the tool developped in Pr. Kirchmair unit (FAME2/FAME3) and initiated the development of a new method to predict xenobiotics metabolism.
- **Germany:** Clémence Frioux visited the lab of Oliver Ebenoh (Heidelberg) for one week.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Selection

10.1.1.1. Member of the Conference Program Committees

- 12th international conference Semantic Web and Tools for Healthcare and Life Sciences (SWAT4HCLS) [O. Dameron]

- 19th NETTAB-BBCC joint international conference [E. Becker, O. Dameron]
- CSBio2019: The 10th International Conference on Computational Systems-Biology and Bioinformatics [A. Siegel]
- WBC@ICML 2019, ICML 2019 Workshop on Computational Biology [A. Siegel]
- CP'2019 25th International Conference on Principles and Practice of Constraint Programming [A. Siegel]
- Journée IA et santé, organisée par l'AFIA (Assoc. Fr. pour l'Intelligence Artificielle) et l'AIM (Assoc. Fr. d'Informatique Médicale) [O. Dameron]
- JOBIM 2019 [A. Siegel]

10.1.1.2. Reviewer

- Semantic Web and Tools for Healthcare and Life Sciences: O. Dameron
- Journée IA et santé: O. Dameron

10.1.1.3. Jury member

- Jury of interpretability prize of Sigmorphon shared task 2019: F. Coste

10.1.2. Journal

10.1.2.1. Member of the Editorial Boards

- Journal of Biomedical Semantics [O. Dameron]
- Editor of special issue in grammatical inference of Machine Learning: [F. Coste]

10.1.2.2. Reviewer - Reviewing Activities

- Bioinformatics [A. Siegel, E. Becker]
- Biosystems [A. Siegel]
- Briefings in Bioinformatics [O. Dameron]
- IEEE/ACM TCBB [A. Siegel]
- Neuroinformatics [O. Dameron]
- Machine Learning [F. Coste]
- PLoS Computational biology [A. Siegel]

10.1.2.3. Peer Community in Genomics

In addition to the traditional publishing activities, Dyliss has also been active in the alternative approach "Peer Community in" (PCI). It is is a non-profit community of researchers who review and recommend for free (we also do it for free with traditional journals and conferences anyway), unpublished preprints in their field (i.e. unpublished articles deposited on open online archives like arXiv.org and bioRxiv.org). To a lesser extent, they may also recommend articles already published in journals.

Denis Tagu has been one of the creators of Peer Community in Genomics (PCI Genomics [More details]), which has been launched in October 2019. The scope of this PCI encompasses all aspects of genomics (structural genomics, functional genomics, epigenomics, evolutionary genomics, population genomics, proteomics, bioinformatics) dealing with every type of organisms (viruses, bacteria, fungi, plants, animals,...) as well as metagenomes.

- Denis Tagu is a manager
- Olivier Dameron is a recommender

10.1.3. Invited Talks

- Dusseldorf, Quantitative Theoretical Biology group, Heinrich Heine Universitat, february 2019 [C. Frioux]
- Colloquium J. Morgenstern (Nice), february 2019 [A. Siegel]
- IGNITE International Training Network Spring school [O. Dameron]
- Colloque d'ouverture 50 ans du Laboratoire Jacques-Louis Lions, Roscoff, march 2019 [A. Siegel]
- CNRS GdR MaDics, june 2019 [O. Dameron]
- Boolean Weekend, Burlington, US, april 2019 [A. Siegel]
- Department of Plant Science, University of Oxford, may 2019 [C. Frioux]
- Quadram Institute, may 2019 [C. Frioux]
- Network analysis to elucidate natural system dynamics, diversity and performance, CECAM Conference, Lyon, may 2019 [A. Siegel]
- Séminaire IMBAS Inra Toulouse, june 2019 [A. Siegel]
- CNRS GdR BIM, november 2019 [O. Dameron]
- The 10th International Conference on Computational Systems-Biology and Bioinformatics (CSBio 2019) [N. Théret] Nice (France) December 4 to 7, 2019

10.1.4. Leadership within the Scientific Community

- Member of the steering committee of the International Conference on Grammatical Inference [F. Coste]

10.1.5. Scientific Expertise

10.1.5.1. International expertise

- Luxembourg, FNR. Core program [A. Siegel]

10.1.5.2. Prospective working groups

- "Big & Open Data en recherche à l'horizon 2040" foresight working group of PROSPER network [F. Coste]

10.1.5.3. National responsibilities

- **Institutional boards for the recruitment and evaluation of researchers**.
  - National Council of Universities, section 65 [O. Dameron, nominated member].
- **Recruitment committees**.
  - Inria Senior Researchers [national committee, A. Siegel]
  - Assistant professor in Mathematics for Biology, section 26, Univ Rennes 1 [E. Becker]

10.1.5.4. National scientific boards

- Animation of the Systems Biology working group of national infrastructure GDR IM and GDR BIM [A. Siegel].
- Board of directors of the French Society for biology of the extracellular matrix [N. Théret].

10.1.5.5. Local responsibilities

- Scientific Advisory Board of Biogenouest [N. Théret]
- Delegate to research integrity at the University of Rennes 1 [N. Théret]
- Head of the "Data and Knowledge Managment" Department (6 teams) of the IRISA lab [A. Siegel, since june 2019]
- IRISA laboratory (computer science department of Univ. Rennes 1) council [A. Siegel, until june 2019]

- Member of the Inria Rennes center council [J. Got]
- Responsability of the IRISA laboratory "Health-biology" cross-cutting axis [O. Dameron, until June 2019] [More details]
- Social commitee of Univ. Rennes 1 [C. Belleannée]
- Scientific committee of Univ. Rennes 1 school of medicine [O. Dameron, A. Siegel].
- Emergency aid commission of Univ. Rennes 1 & Rennes2 [C. Belleannée]

10.1.6. Research Administration

10.1.6.1. Inria Instances

- Inria National evaluation board [A. Siegel, nominated member].
- Equality and diversity Committee, Inria - Responsible of the working group focusing on recruitment procedures [A. Siegel].

10.1.6.2. CNRS

- Chargée de mission "bioinformatique" at INS2I-CNRS (Institute for Computer Science of CNRS) [A. Siegel, nominated member].

10.1.6.3. Inria local instances

- Gender equality commission, IRISA & Inria Rennes [A. Siegel, coordinator]
- CUMI (Comission des utilisateurs des moyens informatiques) of Inria Rennes [F. Coste]
- Inria Rennes PhD recruitment (CORDIs) [C. Belleannée]

10.2. Teaching - Supervision - Juries

10.2.1. Teaching track responsibilities

- Coordination of the doctoral school "Biology and Health" of University of Bretagne Loire, Rennes Site 1 [N. Théret]
- Coordination of the master degree "Bioinformatics", Univ. Rennes1 [E. Becker, O. Dameron]

10.2.2. Course responsibilities

- "Method", Master 2 in Computer Sciences, Univ. Rennes 1 [E. Becker]
- "Statistiques appliquées", 3rd year in Fundamental Computer Sciences, ENS Rennes [E. Becker]
- "Introduction to computational ecology", Master 2 in Ecology, Univ. Rennes 1 [E. Becker]
- "Object oriented programming", Master 1 in Bioinformatics, Univ. Rennes 1 [E. Becker]
- "Advanced R for data analysis", Master 1 in Ecology + Master 1 in Bioinformatics, Univ. Rennes 1 [E. Becker]
- "Insertion Professionnelle et tables rondes", Master 1 and Master 2 in Bioinformatics, Univ. Rennes 1 [E. Becker]
- "Atelier de Biostatistiques", 2nd year Biology, Univ Rennes 1 [E. Becker]
- "Internship", Master 1 in Computer Sciences, Univ. Rennes 1 [C. Belleannée]
- "Supervised machine learning", Master 2 in Computer Sciences, Univ Rennes 1 [F. Coste]
- "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", Master1 in public health, Univ. Rennes 1 [O. Dameron]
- "Programmation en Python", Master 1 in Public Health, Univ. Rennes 1 [O. Dameron]
- "Intégration: Remise à niveau en informatique", Master 1 in bioinformatics, Univ. Rennes 1 [O. Dameron]

- "Programmation impérative en Python", Master 1 in bioinformatics, Univ. Rennes 1 [O. Dameron]
- "Semantic Web and bio-ontologies", Master 2 in bioinformatics, Univ. Rennes 1 [O. Dameron]
- "Internship", Master 2 in bioinformatics, Univ. Rennes 1 [O. Dameron]
- Master: A. Siegel, Integrative and Systems biology, Master 2 in bioinformatics, Univ. Rennes 1 [A. Siegel]
- Micro-environnement Cellulaire normal & pathologique, Master 2 in Biologie cellulaire et Moléculaire, Univ. Rennes 1 [N. Théret]

10.2.3. Teaching

Licence : E. Becker, "TPs Python", 12h, 1st year in Biology, Univ. Rennes 1, France

Licence : E. Becker, "Atelier de Biostatistiques", 34h, 2nd year in Biology, Univ. Rennes 1, France

Licence : E. Becker, "Statistiques Appliquées", 20h, 3rd year in Fundamental Computer Sciences, ENS Rennes, France

Master : E. Becker, "Object oriented programming", 43h, Master 1 in Bioinformatics, Univ. Rennes 1, France

Master : E. Becker, "Advanced R for data analysis", 34h, Master 1 in Bioinformatics, Univ. Rennes 1, France

Master : E. Becker, "Introduction to computational ecology", 34h, Master 2 in Ecology, Univ. Rennes 1, France

Master : E. Becker, "Method", 15h, Master 2 in Computer Sciences, Univ. Rennes 1, France

Master : E. Becker, "Insertion Professionnelle et tables rondes", 8h, Master 1 and Master 2 in Bioinformatics, Univ. Rennes 1, France

Master : E. Becker, "Systems Biology : biological netorks", 27h, Master 2 in Bioinformatics, Univ. Rennes 1, France

Licence: C. Belleannée, Langages formels, 20h, L3 informatique, Univ. Rennes1, France.

Licence: C. Belleannée, Projet professionnel et communication, 16h, L1 informatique, Univ. Rennes1, France.

Licence: C. Belleannée, Enseignant référent, 20h, L1 informatique, Univ. Rennes1, France.

Licence: C. Belleannée, Spécialité informatique : Functional and immutable programming , 42h, L1 informatique, Univ. Rennes1, France

Master: C. Belleannée, Algorithmique du texte et bioinformatique, 10h, M1 informatique, Univ. Rennes1, France

Master: S. Blanquart, Juries of Master 1 in bioinformatics, 5h, Univ. Rennes 1, France

Master: L. Bourneuf, Projet, 25h, M1 Santé Publique, France.

Master: F. Coste, Supervised machine learning, 20h, M2 + M1 Science Informatique, Univ. Rennes, France

Licence: O. Dameron, Biostatistiques, 12h, 1st year school of medicine, Univ. Rennes 1, France

Licence: O. Dameron, "Atelier bioinformatique", 12h, Licence 2 informatique, Univ. Rennes 1, France

Master: O. Dameron, "Intégration: Remise à niveau en informatique", 18h Master 1 in bioinformatics, Univ. Rennes 1, France

Master: O. Dameron, "Programmation impérative en Python", 985h Master 1 in bioinformatics, Univ. Rennes 1, France

Master: O. Dameron, "Système informatique GNU/Linux", 10h, Master 1 in bioinformatics, Univ. Rennes 1, France

Master: O. Dameron, 5h, "Internship", Master 1 in bioinformatics, Univ. Rennes 1, France

Master: O. Dameron, "Bases de mathématiques et probablité", 9h, Master1 in public health, Univ. Rennes 1, France

Master: O. Dameron, "Programmation impérative en Python (2)", 3h Master 1 in public health, Univ. Rennes 1, France

Master: O. Dameron, 20h, "Semantic Web and bio-ontologies", Master 2 in bioinformatics, Univ. Rennes 1, France

Master: O. Dameron, 15h, "Internship", Master 2 in bioinformatics, Univ. Rennes 1, France

Licence: N. Guillaudeux, Projet professionnel et communication, 16h, 1st year Computer Science, Univ. Rennes 1, France

Licence: N. Guillaudeux, "TPs Python", 36h, 1st year in Biology, Univ. Rennes 1, France

Licence: M. Louarn, Introduction à la BioInformatique, 6h, L2 Informatique, Univ. Rennes 1, France.

Licence: M. Louarn, Informatique, 10h, L1 Physique Chimie, Univ. Rennes 1, France.

Master: M. Louarn, Informatique Médicale Avancée, 2h, M1 Médecine, Univ. Rennes 1, France.

Master: M. Louarn, Object-oriented programming, 25h, M2 bioinformatique et génomique, Univ. Rennes 1, France.

Master: M. Louarn, Jury de stage, 6h, M1 bioinformatique et génomique, Univ. Rennes 1, France.

10.2.4. Supervision

PhD: Yann Rivault, *Analyse de trajectoires de soins à partir de bases de données médico-administratives : apport d'un enrichissement pas des connaissances biomédicales issues du Web des Données*, defended 28th January 2019, supervised by O. Dameron and N. Lemeur [64].

PhD: Lucas Bourneuf, *A search space of graph motifs for graph compression: From Powergraphs to triplet concepts*, defended 17th december 2019, supervised by J. Nicolas [12].

PhD in progress: Johanne Bakalara, *Temporal models of care sequences for the exploration of medico-administrative data*, started in Oct. 2018, supervised by T. Guyet (Lacodam), E. Oger (Repères) and O. Dameron.

PhD in progress: Arnaud Belcour, *Inferring Model metabolisms for bacterial ecosystems reduction*, started in Oct. 2019, supervised by A. Siegel and S. Blanquart.

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: Olivier Dennler, *Modular functional characterization of ADAMTL and ADAMTSL protein families*, started in Oct. 2019, supervised by N. Theret, F. Coste and S. Blanquart.

PhD in progress: Nicolas Guillaudeux, *Compare gene structures to predict isoform transcripts*, started in Oct. 2018, supervised by O. Dameron, S. Blanquart and C. Belleannée.

PhD in progress: Virgilio Kmetzsch *Multi-modal analysis of neuroimaging and transcriptomics data in genetically-induced fronto-temporal dementia*, started in Oct. 2019, supervised by E. Becker and O. Colliot (Inria Aramis, ICM Paris)

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 : Hugo Talibart, Learning grammars with long-distance correlations on proteins, started in Nov. 2017, supervised by F. Coste and J. Nicolas.

PhD in progress : Pierre Vignet, *Identification et conception expérimentale de nouveaux agents thérapeutiques à partir d'un modèle informatique des réseaux d'influence du TGF-β dans les pathologies hépatiques chroniques* , started in Dec. 2018, supervised by N. Théret and A. Siegel.

PhD in progress : Méline Wéry, *Methodology development in disease treatment projects.* , started in Oct. 2017, supervised by E. Becker, C. Bettembourg (Sanofi), O. Dameron, and A. Siegel.

10.2.5. Juries

- Referee of PhD thesis: Aarón Ayllón-Benítez, Univ. Bordeaux [O. Dameron] - A. Beica, ENS [A. Siegel, president] - F. Bridoux, Marseille [A. Siegel] - C. Hernandez, ENS [A. Siegel, president]
- Member of PhD thesis juries: Sandeep Manandhar, Univ. Rennes [O. Dameron] - A. Husson, Paris Diderot [A. Siegel] - Arran Hodgkinson, Univ. Montpellier [N. Théret]
- Referee of habilitation thesis: Fleur Mougin, Univ. Bordeaux [O. Dameron] - S. Peres, Orsay [A. Siegel]
- Member of habilitation thesis juries: Adrien Coulet, Univ. Lorraine [O. Dameron] - L. Calzone [A. Siegel, president]

10.2.6. Interns

- Internship, from Jan 2019 until Jun 2019. Supervised by O. Dameron and O. Corby (Wimmics team, Sophia-Antipolis) . Student: Antoine Abel (Master 2 bioinformatique, Univ. Rennes 1). Subject: Faster SPARQL federated queries [31]
- Internship, from Jan 2019 until Jun 2019, Supervised by O. Dameron and A. Duchene (Theranexus). Student: Pierre Beaudier (Master 2 bioinformatique, Rennes). Subject: Evaluation of the public databases' relevance as comprehensive tools for pharmalogical mechanisms [32]
- Internship, from Jan 2019 until Jun 2019, Supervised by E. Becker, Fabrice Legeai (IGEPP) and Julien Bobe (LPGP). Student: Fanny Casse (Master 2 bioinformatique, Rennes). Subject: MiRNA regulating genes responsible for fertility and early embryonic development in the Medaka fish [35].
- Internship, from April 2019 until July 2019, Supervised by P. Dabert (IRSTEA), A. Siegel and S. Blanquart. Student: Theo Combe (Master 1 bioinformatique, Rennes). Subject: Reconstructing metabolic pathways of methanogenic communities using metabarcoding data
- Internship, from May 2019 until Jun 2019, Supervised by F. Coste. Student: Arthur Correnson (Licence math-informatique, Université Paul Sabatier, Toulouse). Subject: Implementation of a new partial local multiple alignment algorithm.
- Internship, from Jan 2019 until Jun 2019, Supervised by F. Coste, N. Théret, S. Blanquart, C. Belleannée. Student: Olivier Dennler (Master 2 bioinformatique, Rennes). Subject: Functional module characterization of ADAMTS / ADAMTSL protein family [37]
- Internship, from Jan 2019 until Jun 2019, Supervised by A. Siegel. Student: Clara Emery (Master 2 bioinformatique, Rennes). Subject: Modeling of quorum sensing system by in silico approaches
- Internship, from May 2019 until Oct 2019, Supervised by E. Becker and G. Rabut (IGDR, Rennes). Student: Camille Juigne (ENSSAT, Lannion). Subject: Data integration and formal analysis of the network between ubiquitination enzymes.
- Internship, from Jan 2019 until Jun 2019, Supervised by F. Morrews. Student: Albane Lysiak (Master 2 bioinformatique, Rennes). Subject: Exploiting paralog genes for discovering genetic regulation by transcription factors.
- Internship, from Jul 2019 until Aug 2019, Supervised by C. Belleannée and N. Guillaudeux. Student: Pierre Alexis Odye (Master 1 informatique, Rennes). Subject: Design of a database on known and predicted transcriptomics data in humans, mice and dogs.
- Internship, from Jul 2019 until Aug 2019, Supervised by E. Becker and F. Moreews. Student: Hugo Simon (1st year Telecom Paris Tech). Subject: Large scale analysis of metabolic networks as biological graphs.
- Internship, from Jun 2019 until Jul 2019, Supervised by A. Siegel and R. Andonov. Student: Kerian Thuillier (ENS Rennes). Subject: Linear programming for metabolic network completion [41]

- Internship, from Jan 2019 until Jul 2019, Supervised by C. Frioux. Student: Margot Wagner (Master 2 écologie, Rennes). Subject: Discriminant analysis of transcriptomics and metagenomics of an alga and its microbiota in multiple culture conditions with genome-scale modelling.

10.3. Popularization

10.3.1. Interventions

- "J'peux pas j'ai informatique", Welcoming of high-school students to leverage stereotypes about computer sciences (110 participants), Apr. 2019 [O. Dameron, A. Siegel]
- "Elles codent Elles créent" Our Ph-D students have been involved in creative Python learning sessions for female students in two high-schools. [M. Louarn, M. Wéry, A. Siegel]
- Operation DECLIC (Dialogues Entre Chercheurs et Lycéens pour les Intéresser à la Construction des Savoirs). Lycée Descartes, Rennes. Nov 2019 [More details] [N. Théret, M. Conan, P. Vignet, O. Dennler]

10.3.2. Internal action

- Organization of the weekly seminar "Symbiose", involving teams DYLISS, GENSCALE and GENOUEST: [S. Blanquart]

10.3.3. Creation of media or tools for science outreach

Science en Cour[t]s (*http://sciences-en-courts.fr/*) Many of our current and former PhD students (M.Wéry, L. Bourneuf, 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 PhD. 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 [More details]

11. Bibliography

Major publications by the team in recent years

[1] M. Aite, M. Chevallier, C. Frioux, C. Trottier, J. Got, M.-P. Cortés, S. N. Mendoza, G. Carrier, O. Dameron, N. Guillaudeux, M. Latorre, N. Loira, G. V. Markov, A. Maass, A. Siegel. *Traceability, reproducibility and wiki-exploration for "à-la-carte" reconstructions of genome-scale metabolic models*, in "PLoS Computational Biology", May 2018, vol. 14, n^{o} 5, e1006146 [*DOI :* 10.1371/journal.pcbi.1006146], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01807842

[2] 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

[3] 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

[4] 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^{o} 1, p. 106-117 [*DOI :* 10.1002/mbo3.315], https://hal.inria.fr/hal-01246173

[5] 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, Lecture Notes in Computer Sciences, September 2017, 17, https://hal.archives-ouvertes.fr/hal-01559249

[6] 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

[7] F. COSTE, J. NICOLAS.*Learning local substitutable context-free languages from positive examples in polynomial time and data by reduction*, in "ICGI 2018 - 14th International Conference on Grammatical Inference", Wrocław, Poland, September 2018, vol. 93, p. 155 - 168, https://hal.inria.fr/hal-01872266

[8] C. FRIOUX, E. FREMY, C. TROTTIER, A. SIEGEL.*Scalable and exhaustive screening of metabolic functions carried out by microbial consortia*, in "Bioinformatics", September 2018, vol. 34, n^{o} 17, p. i934 - i943 [*DOI :* 10.1093/BIOINFORMATICS/BTY588], https://hal.inria.fr/hal-01871600

[9] C. FRIOUX, T. SCHAUB, S. SCHELLHORN, A. SIEGEL, P. WANKO.*Hybrid Metitebolic Network Completion*, in "Theory and Practice of Logic Programming", November 2018, p. 1-23, https://hal.inria.fr/hal-01936778

[10] 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^{o} 1, 32 [*DOI :* 10.1371/JOURNAL.PCBI.1005276], https://hal.inria.fr/hal-01449100

[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] L. BOURNEUF.*A search space of graph motifs for graph compression : From Powergraphs to triplet concepts*, Université Rennes 1, December 2019, https://hal.inria.fr/tel-02399641

Articles in International Peer-Reviewed Journal

[13] S. BLANQUART, A.-S. BOROWIEC, P. DELCOURT, M. FIGEAC, C. A. EMERLING, A. S. MESEGUER, M. ROUDBARAKI, N. PREVARSKAYA, G. BIDAUX.*Evolution of the human cold/menthol receptor, TRPM8*, in "Molecular Phylogenetics and Evolution", July 2019, vol. 136, p. 104-118 [*DOI :* 10.1016/J.YMPEV.2019.04.011], https://hal.inria.fr/hal-02284919

[14] V. DELANNÉE, S. LANGOUËT, A. SIEGEL, N. THÉRET.*In silico prediction of Heterocyclic Aromatic Amines metabolism susceptible to form DNA adducts in humans*, in "Toxicology Letters", January 2019, vol. 300, p. 18-30 [*DOI :* 10.1016/J.TOXLET.2018.10.011], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01903264

[15] W. DYRKA, M. PYZIK, F. COSTE, H. TALIBART.*Estimating probabilistic context-free grammars for proteins using contact map constraints*, in "PeerJ", March 2019, vol. 7, p. 1-35 [*DOI :* 10.7717/PEERJ.6559], https://hal.archives-ouvertes.fr/hal-02400871

[16] M. FOLSCHETTE, V. LEGAGNEUX, A. PORET, L. CHEBOUBA, C. GUZIOLOWSKI, N. THÉRET.*A pipeline to create predictive functional networks: application to the tumor progression of hepatocellular carcinoma*, in "BMC Bioinformatics", January 2020 [*DOI :* 10.1186/S12859-019-3316-1], https://hal.archives-ouvertes.fr/hal-02095930

[17] D. NÈGRE, M. AITE, A. BELCOUR, C. FRIOUX, L. BRILLET-GUÉGUEN, X. LIU, P. BORDRON, O. GODFROY, A. P. LIPINSKA, C. LEBLANC, A. SIEGEL, S. M. DITTAMI, E. CORRE, G. V. MARKOV.*Genome–Scale Metabolic Networks Shed Light on the Carotenoid Biosynthesis Pathway in the Brown Algae Saccharina japonica and Cladosiphon okamuranus*, in "Antioxidants", November 2019, vol. 8, n^o 11, 564 [*DOI :* 10.3390/ANTIOX8110564], https://hal.inria.fr/hal-02395080

[18] Y. RIVAULT, O. DAMERON, N. LE MEUR.*queryMed: Semantic Web functions for linking pharmacological and medical knowledge to data*, in "Bioinformatics", January 2019, vol. 35, n^o 17, p. 3203-3205 [*DOI :* 10.1093/BIOINFORMATICS/BTZ034], https://hal.archives-ouvertes.fr/hal-01988699

[19] K. STRAUB, M. LINDE, C. KROPP, S. BLANQUART, P. BABINGER, R. MERKL.*Sequence selection by FitSS4ASR alleviates ancestral sequence reconstruction as exemplified for geranylgeranylglyceryl phosphate synthase*, in "Biological Chemistry", February 2019, vol. 400, n^o 3, p. 367-381 [*DOI :* 10.1515/HSZ-2018-0344], https://hal.inria.fr/hal-02284913

[20] M. WERY, O. DAMERON, J. NICOLAS, E. RÉMY, A. SIEGEL.*Formalizing and enriching phenotype signatures using Boolean networks*, in "Journal of Theoretical Biology", 2019, vol. 467, p. 66-79 [*DOI :* 10.1016/J.JTBI.2019.01.015], https://hal.inria.fr/hal-02018724

Invited Conferences

[21] A. SIEGEL.*Learning boolean regulations based on prior-knowledge data: a logical-based viewpoint*, in "SASB 2019 - 10th International Workshop on Static Analysis for Systems Biology", Porto, Portugal, October 2019, https://hal.inria.fr/hal-02412421

[22] A. SIEGEL.*Using automated reasoning to explore unconventional organisms: a first step to explore host-microbial interactions*, in "MPA 2019 - Conference on Metabolic Pathway Analysis", Riga, Latvia, August 2019, https://hal.inria.fr/hal-02412419

International Conferences with Proceedings

[23] J. BAKALARA, T. GUYET, O. DAMERON, E. OGER, A. HAPPE.*Temporal models of care sequences for the exploration of medico-administrative data*, in "2019 - Workshop IA&Santé, PFIA", Toulouse, France, July 2019, p. 1-8, https://hal.archives-ouvertes.fr/hal-02265743

[24] L. BOURNEUF, J. NICOLAS.*Concept Lattices as a Search Space for Graph Compression*, in "ICFCA 2019 - 15th International Conference on Formal Concept Analysis", Francfort, Germany, D. C. L. B. SERTKAYA (editor), ICFCA: International Conference on Formal Concept Analysis, Springer, May 2019, vol. 15th International Conference, n^o 15, p. 274-289 [*DOI :* 10.1007/978-3-030-21462-3_18], https://hal.inria.fr/hal-02399578

[25] M. LOUARN, F. CHATONNET, X. GARNIER, T. FEST, A. SIEGEL, O. DAMERON.*Increasing life science resources re-usability using Semantic Web technologies*, in "eScience 2019 - 15th International eScience Conference", San Diego, United States, September 2019, p. 1-9, https://hal.inria.fr/hal-02274982

Conferences without Proceedings

[26] L. BOURNEUF.*Biseau: An Answer Set Programming Environment for High-Level Specification and Graph Visualization applied to FCA*, in "ICFCA 2019 - International Conference on Formal Concept Analysis", Frankfort, Germany, June 2019, p. 1-6, https://hal.inria.fr/hal-02399610

[27] X. GARNIER, A. BRETAUDEAU, F. LEGEAI, A. SIEGEL, O. DAMERON.*AskOmics: a user-friendly interface to Semantic Web technologies for integrating local datasets with reference resources*, in "JOBIM 2019 - Journées Ouvertes Biologie, Informatique et Mathématiques", Nantes, France, July 2019, 1, https://hal.archives-ouvertes.fr/hal-02401750

[28] V. J. HENRY, G. BASSIGNANA, V. ZUJOVIC, F. DE VICO FALLANI, O. DAMERON, I. MOSZER, O. COLLIOT.*Conciliation of process description and molecular interaction networks using logical properties of ontology*, in "JOBIM 2019 - Journées Ouvertes Biologie, Informatique et Mathématiques", Nantes, France, July 2019, https://hal.archives-ouvertes.fr/hal-02301702

[29] N. LE MEUR, Y. RIVAULT, O. DAMERON.*Linking pharmacological and medical knowledge using semantic Web technologies*, in "useR! 2019 - Conférence internationale des utilisateurs de R", Toulouse, France, July 2019, https://hal.archives-ouvertes.fr/hal-02294364

[30] H. TALIBART, F. COSTE.*Using residues coevolution to search for protein homologs through alignment of Potts models*, in "CECAM 2019 - workshop on Co-evolutionary methods for the prediction and design of protein structure and interactions", Lausanne, Switzerland, CECAM-HQ-EPFL, June 2019, p. 1-2, https://hal.inria.fr/hal-02402646

Other Publications

[31] A. ABEL.*Faster SPARQL Federated Queries*, Université Rennes1, June 2019, https://hal.inria.fr/hal-02269417

[32] P. BEAUDIER.*Evaluation of the public databases' relevance as comprehensive tools for pharmacological mechanisms*, Université de Rennes 1, June 2019, https://hal.inria.fr/hal-02191210

[33] A. BELCOUR, C. FRIOUX, M. AITE, A. BRETAUDEAU, A. SIEGEL.*Metage2Metabo: metabolic complementarity applied to genomes of large-scale microbiotas for the identification of keystone species*, December 2019, working paper or preprint [*DOI :* 10.1101/803056], https://hal.inria.fr/hal-02395024

[34] B. BURGUNTER-DELAMARE, H. KLEINJAN, C. FRIOUX, E. FREMY, M. WAGNER, E. CORRE, A. LE SALVER, C. LEROUX, C. LEBLANC, A. SIEGEL, S. M. DITTAMI.*Metabolic complementarity between a brown alga and associated 1 cultivable bacteria provide indications of beneficial interactions*, October 2019, working paper or preprint [*DOI :* 10.1101/813683], https://hal.sorbonne-universite.fr/hal-02333039

[35] F. CASSE ENCADRÉE.*Regulation of genes regulating fecondity and early embryonic development by miRNAs in the Medaka fish*, Univ Rennes, Inria, CNRS, IRISA, June 2019, https://hal.inria.fr/hal-02192476

[36] F. COSTE.*Deep learning languages: a key fundamental shift from probabilities to weights?*, August 2019, https://arxiv.org/abs/1908.00785 - working paper or preprint, https://hal.inria.fr/hal-02235207

[37] O. DENNLER.*Caractérisation en modules fonctionnels de la famille de protéines ADAMTS / ADAMTSL*, Univ Rennes, June 2019, https://hal.inria.fr/hal-02403084

[38] S. M. DITTAMI, E. CORRE, L. BRILLET-GUÉGUEN, A. P. LIPINSKA, M. AITE, K. AVIA, C. CARON, C. H. CHO, J. COLLEN, A. CORMIER, L. DELAGE, S. DOUBLEAU, C. FRIOUX, A. GOBET, I. GONZÁLEZ-NAVARRETE, A. GROISILLIER, C. HERVÉ, D. JOLLIVET, H. KLEINJAN, C. LEBLANC, X. LIU, D. MARIE, G. V. MARKOV, A. E. MINOCHE, P. PERICARD, M.-M. PERRINEAU, A. F. PETERS, A. SIEGEL, C. TROTTIER, H. S. YOON, H. HIMMELBAUER, C. BOYEN, T. TONON.*The genome of Ectocarpus subulatus -a highly stress-tolerant brown alga*, October 2019, working paper or preprint [*DOI :* 10.1101/307165], https://hal.sorbonne-universite.fr/hal-02333021

[39] N. GUILLAUDEUX, C. BELLEANNÉE, S. BLANQUART, J.-S. VARRÉ.*Predicting isoform transcripts: What does the comparison of known transcripts in human, mouse and dog tell us?*, July 2019, vol. 8, 1, JOBIM 2019 - Journées Ouvertes Biologie, Informatique et Mathématiques, Poster présenté à JOBIM 2019 (https://jobim2019.sciencesconf.org/) qui s'est déroulé du 2 au 5 juillet 2019 et à ISMB/ECCB 2019 (https://www.iscb.org/ismbeccb2019) qui s'est déroulé du 21 au 25 juillet 2019 [*DOI :* 10.7490/F1000RESEARCH.1117311.1], https://hal.inria.fr/hal-02267357

[40] H. TALIBART, F. COSTE.*Using residues coevolution to search for protein homologs through alignment of Potts models*, July 2019, 1, JOBIM 2019 - Journées Ouvertes Biologie, Informatique et Mathématiques, Poster, https://hal.inria.fr/hal-02402687

[41] K. THUILLIER.*Linear programming for metabolic network completion*, Inria Rennes - Bretagne Atlantique and University of Rennes 1, France, December 2019, https://hal.inria.fr/hal-02408003

[42] M. WERY, E. BECKER, F. AUGE, C. BETTEMBOURG, O. DAMERON, A. SIEGEL.*Identification of causal signature using omics data integration and network reasoning-based analysis*, July 2019, 1, JOBIM 2019 - Journées Ouvertes Biologie, Informatique et Mathématiques, Poster, https://hal.inria.fr/hal-02193860

References in notes

[43] 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^o^ 1, 30 [*DOI :* 10.1186/1752-0509-8-30], http://www.hal.inserm.fr/inserm-00978313

[44] T. BERNERS LEE, W. HALL, J. A. HENDLER, K. O'HARA, N. SHADBOLT, D. J. WEITZNER.*A Framework for Web Science*, in "Foundations and Trends in Web Science", 2007, vol. 1, n^o^ 1, p. 1–130

[45] 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^o^ 1, e86525 [*DOI :* 10.1371/JOURNAL.PONE.0086525], https://hal.inria.fr/hal-00941850

[46] S. BLANQUART, J.-S. VARRÉ, P. GUERTIN, A. PERRIN, A. BERGERON, K. M. SWENSON.*Assisted transcriptome reconstruction and splicing orthology*, in "BMC Genomics", Nov 2016, vol. 17, n^o^ 10, 786, https://doi.org/10.1186/s12864-016-3103-6

[47] 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^o^ 1, 32 [*DOI :* 10.1186/1752-0509-8-32], https://hal.inria.fr/hal-00980499

[48] 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^o 1, p. 106-117 [*DOI :* 10.1002/MBO3.315], https://hal.inria.fr/hal-01246173

[49] M. Boutet, L. Gauthier, M. Leclerc, G. Gros, V. De Montpreville, N. Théret, E. Donnadieu, F. Mami-Chouaib. *TGF-β signaling intersects with CD103 integrin signaling to promote T lymphocyte accumulation and antitumor activity in the lung tumor microenvironment*, in "Cancer Research", 2016, vol. 76, n^o 7, p. 1757-69 [*DOI :* 10.1158/0008-5472.CAN-15-1545], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01282442

[50] 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

[51] 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

[52] 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

[53] 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

[54] V. Delannée, S. Langouët, N. Théret, A. Siegel. *A 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

[55] 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

[56] K. Faust, J. Raes. *Microbial interactions: from networks to models*, in "Nat. Rev. Microbiol.", Jul 2012, vol. 10, n^o 8, p. 538–550

[57] M. Y. Galperin, D. J. Rigden, X. M. Fernández-Suárez. *The 2015 nucleic acids research database issue and molecular biology database collection*, in "Nucleic acids research", 2015, vol. 43, n^o D1, p. D1–D5

[58] 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

[59] 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

[60] 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^o 5, e96491 [*DOI :* 10.1371/JOURNAL.PONE.0096491], https://hal.inria.fr/hal-00989635

[61] M.-A. LAURENT, D. BONNIER, N. THÉRET, P. TUFFÉRY, G. MOROY.*In silico characterization of the interaction between LSKL peptide, a LAP-TGF-beta derived peptide, and ADAMTS1*, in "Computational Biology and Chemistry", April 2016, vol. 61, p. 155-161 [*DOI :* 10.1016/J.COMPBIOLCHEM.2016.01.012], https://hal.archives-ouvertes.fr/hal-02394687

[62] 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

[63] 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^o 1, 32 [*DOI :* 10.1371/JOURNAL.PCBI.1005276], https://hal.inria.fr/hal-01449100

[64] Y. RIVAULT.*Care trajectory analysis using medico-administrative data : contribution of a knowledge-based enrichment from the Linked Data*, Université Rennes 1, January 2019, https://tel.archives-ouvertes.fr/tel-02137442

[65] M. H. SAIER, V. S. REDDY, B. V. TSU, M. S. AHMED, C. LI, G. MORENO-HAGELSIEB.*The Transporter Classification Database (TCDB): recent advances*, in "Nucleic Acids Res.", Jan 2016, vol. 44, n^o D1, p. D372–379

[66] D. B. SEARLS.*String variable grammar: A logic grammar formalism for the biological language of DNA*, in "The Journal of Logic Programming", 1995, vol. 24, n^o 1, p. 73 - 102, Computational Linguistics and Logic Programming [*DOI :* 10.1016/0743-1066(95)00034-H], http://www.sciencedirect.com/science/article/pii/074310669500034H

[67] 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^o 7, e1002195

[68] R. VERBORGH, M. VANDER SANDE, O. HARTIG, J. VAN HERWEGEN, L. DE VOCHT, B. DE MEESTER, G. HAESENDONCK, P. COLPAERT.*Triple Pattern Fragments: a Low-cost Knowledge Graph Interface for the Web*, in "Journal of Web Semantics", March 2016, vol. 37–38, p. 184–206 [*DOI :* DOI:10.1016/J.WEBSEM.2016.03.003], http://linkeddatafragments.org/publications/jws2016.pdf

Project-Team EASE

Enabling Affordable Smarter Environment

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
Distributed programming and Software engineering

Table of contents

Project-Team EASE

Creation of the Team: 2018 January 01, updated into Project-Team: 2019 March 01

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
A2.5.1. - Software Architecture & Design
A2.5.3. - Empirical Software Engineering
A2.5.4. - Software Maintenance & Evolution
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
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 city

B8.5.2. - Crowd sourcing
B8.5.3. - Collaborative economy
B9.8. - Reproducibility
B9.10. - Privacy

1. Team, Visitors, External Collaborators

Research Scientists
Paul Couderc [Inria, Researcher]
Christophe Couturier [IMT Atlantique, Researcher]

Faculty Members
Jean-Marie Bonnin [Team leader, IMT Atlantique, Professor, HDR]
Yoann Maurel [Univ de Rennes I, Associate Professor]
Frédéric Weis [Univ de Rennes I, Associate Professor, HDR]

Technical Staff
Elisa Blanchard [IMT Atlantique, Engineer, from Dec 2019]
Alexis Girard [Inria, Engineer, from Nov 2019]
Nicolas Harel [Inria, Engineer, from Dec 2019]
Laudin Molina [IMT Atlantique, Engineer, from Apr 2019]

PhD Students
Indra Ngurah [IMT Atlantique, PhD Student]
Alexandre Rio [OKWIND, PhD Student, granted by CIFRE]
Rodrigo Silva [IMT Atlantique, PhD Student, until Mar 2019]

Visiting Scientist
Ikram Cherifi [IMT Atlantique, until Feb 2019]

Administrative Assistant
Fabienne Cuyollaa [Inria, Administrative Assistant]

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 remain 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/EASE 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 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).

EASE 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).

3. Research Program

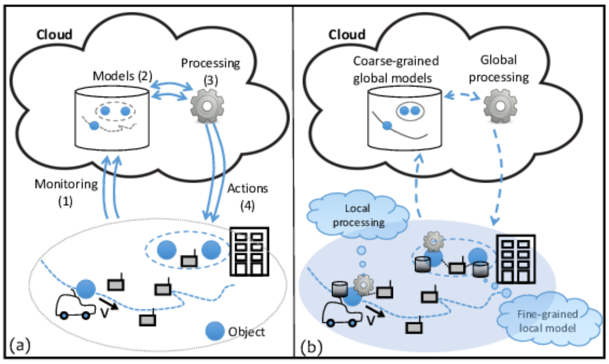


Figure 1. Adaptation processes in pervasive environments

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 systems. 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 crossing 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 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 EASE 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 really 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 aggregates it. These could lead to a modification of the behavior of the network layer and the path the piece of data uses in the aggregation process.

Acting on object interactions. Objects in the environment could adapt their behavior in a way that strongly depends 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 extend our previous work 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*. EASE 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. Automation in Smart City

The domain of Smart Cities is still young but it is already a huge market which attracts 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 vehicles 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 is 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. That helps 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 example of smart city automated processes that will have to share information within the surrounding to make their decisions.

4.3. 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

KEYWORD: Contextual service

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

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).
KEYWORDS: Composite objects - RFID

- Participants: Paul Couderc and Alexis Girard (Univ. Rennes 1)
- Partner: Univ. Rennes 1 (IETR - lab bringing together researchers in the electronics and telecommunications)
- Contact: Paul Couderc

5.3. ISO/IEC 15118-2 Open source Implementation

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.

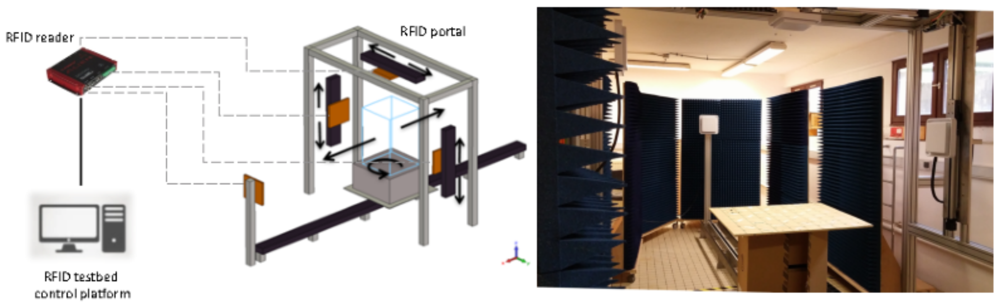


Figure 2. RFID testbed

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 [0].

KEYWORDS: Smart Grid - Intelligent Transport System

- Partner: IMT Atlantique
- Contact: Jean-Marie Bonnin

6. New Results

6.1. Smart City and ITS

Participants: Indra Ngurah, Christophe Couturier, Rodrigo Silva, Frédéric Weis, Jean-Marie Bonnin [contact].

In the last years, we contributed to the specification of the hybrid (ITS-G5 + Cellular) communication architecture of the French field operation test project SCOOP@F. The proposed solution relies on the MobileIP family of standards and the ISO/ETSI ITS Station architecture we contributed to standardize at IETF and ISO. On this topic our contribution mainly focussed on bringing concepts from the state of the art to real equipments. For the last year of the SCOOP@F part 2 project, we took part to the performance evaluation process by providing a test and validation platform for IP mobility protocols (MobileIP, NEMO) and IPsec cyphering. This platform allows us to identify the performance limits of current implementation of mobility and security protocols. Moreover it spotted implementation incompatibilities between the open source implementations of theses protocols (namely UMIP and StrongSwan) and helped the industrial partners of the project to identify associated risks.

InDiD is the logical follow up of SCOOP@F part 2. This 3.5 years long European project (mid 2019-2023) aims at testing ITS applications on a large scale national deployment of connected vehicles and infrastructure. This version of the project specifically complex use cases (so called day 1.5) and urban application. For the beginning of this project, we proposed several innovative use cases. Our "Backward cartography update" scenario has been selected as a priority candidate for implementation. In line with the collaborative approaches of EASE, we propose to use vehicles' observations to inform other vehicles and/or a cartography server about differences between the digital map and the reality.

We also want to explore the benefits of new capabilities of upcoming communication technologies to enrich the interactions between vehicle and smartphones or wearable devices. We defined an architecture for both localisation and communication with vulnerable users (workers in road and construction works). Short range communications between dangers (maneuvering construction vehicles) and workers rely on the advertisement feature of Bluetooth Low Energy (BLE). This connectionless communication mode enables for easy direct communication between any node in the neighborhood. It is inspired from the ITS-Station communication standard and we aim to integrate our work into future versions of the standards. Another contribution in this project aims at enhancing the localisation precision in harsh conditions. Recent version of radio communication standards (eg. Bluetooth 5.1 or 802.11ax) now integrate intrinsic real time localisation primitives giving information such as Angle of Arrival (AoA), Angle of Departure (AoD) or distance evaluation based on Time of Flight (ToF) measurements. We started to study how to merge this information with other localisation evidence sources and how to structure a collaborative framework to share it with other objects in the environment. This early works opens the doors to many other works in the future.

[0] http://www.vedecom.fr/

The development of innovative applications for smart cities has also been made possible by the rise of Internet of Things and especially the deployment of numerous low energy devices. The collection of the huge amount of data produced by all these piece of hardware become a challenge for the communication networks. In smart cities, the mobility of vehicles can be used to collect data produced by connected objects and to deliver them to several applications which are delay tolerant. The Vehicular Delay Tolerant Networks (VDTN) can be utilized for such services. We designed DC4LED (Data Collection for Low Energy Devices): a hierarchical VDTN routing which takes advantage of the specific mobility patterns of the various type of vehicles. It provides a low-cost delivery service for applications that need to gather data generated from the field. The idea is to propose a simple routing scheme where cars, taxis, and buses route data hierarchically in a store-carry-forward mechanism to any of the available Internet Point-of-Presences in the city. We compare using simulation tools the performance of DC4LED routing with two legacy VDTN routing schemes which represent the extreme ends of VDTN routing spectrum: First-contact and Epidemic routing. It show that DC4LED has much lower network overhead in comparison with the two legacy routing schemes, which is advantageous for its implementation scalability. The DC4LED also maintains comparable data delivery probability and latency to Epidemic routing.

The situational viewing and surveillance in cities is one such category of applications which can benefit from various networking solutions available to transport images or data from installed sensor cameras. We explore how our DC4LED mechanism can be used to for a city-wide image and data collection service. We study the networking performance in terms of increasing image sizes that can be transported with respect to varying vehicular density in city. We focus mainly on two technologies for sensors to vehicles communications: ZigBee and ITS-G5. We show that, surprisingly such very simple mechanism could meet the requirements of multiple services.

6.2. Autonomic Maintenance of Optical Networks

Participant: Jean-Marie Bonnin [contact].

The application of classification techniques based on machine learning approaches to analyze the behavior of network users has interested many researchers in the last years. In a recent work, we have proposed an architecture for optimizing the upstream bandwidth allocation in Passive Optical Network (PON) based on the traffic pattern of each user. Clustering analysis was used in association with an assignment index calculation in order to specify for PON users their upstream data transmission tendency. A dynamic adjustment of Service Level Agreement (SLA) parameters is then performed to maximize the overall customers' satisfaction with the network. In this work, we extend the proposed architecture by adding a prediction module as a complementary to the first classification phase. Grey Model GM(1,1) is used in this context to learn more about the traffic trend of users and improve their assignment. An experimental study is conducted to show the impact of the forecaster and how it can overcome the limits of the initial model.

This work has been done in collaboration with IRISA-OCIF team.

6.3. Location assessment from local observations

Participants: Yoann Maurel, Paul Couderc [contact].

Confidence in location is increasingly important in many applications, in particular for crowd-sensing systems integrating user contributed data/reports, and in augmented reality games. In this context, some users can have an interest in lying about their location, and this assumption has been ignored in several widely used geolocation systems because usually, location is provided by the user's device to enhance the user's experience. Two well known examples of applications vulnerable to location cheating are Pokemon Go and Waze.

Unfortunately, location reporting methods implemented in existing services are weakly protected: it is often possible to lie in simple cases or to emit signals that deceive the more cautious systems. For example, we have experimented simple and successful replay attacks against Google Location using this approach, as shown on Figure 3.

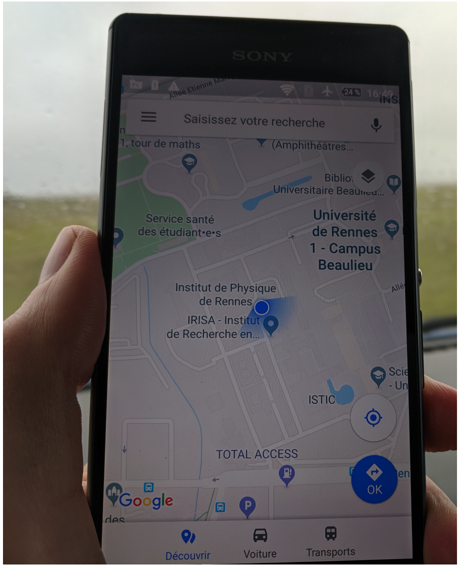


Figure 3. Google map deceived by faked Wi-Fi beacons replayed by an ESP-32

An interesting idea consists in requiring user devices to prove their location, by forcing a secure interaction with a local resource. This idea has been proposed by several works in the literature; unfortunately, this approach requires ad hoc deployment of specific devices in locations that are to be "provable".

We proposed an alternative solution using passive monitoring of Wi-Fi traffic from existing routers. The principle is to collect beacon timestamp observations (from routers) and other attributes to build a knowledge that requires frequent updates to remains valid, and to use statistical test to validate further observations sent by users. Typically, older data collected by a potential attacker will allow him to guess the current state of the older location for a limited timeframe, while the location validation server will get updates allowing him to determine a probability of cheating request.The main strength is its ability to work on existing Wi- Fi infrastructures, without specific hardware. Although it does not offer absolute proof, it makes attacks much more challenging and is simple to implement.

This work was published at CCNC'2019 [1]. We are currently working in broadening this approach, in particular using other attributes of Wi-Fi traffic beside beacon timestamps, and combining the timestamp solution with other type of challenges to propose a diversity of challenges for location validation servers. We are also working on the attack side, which presents interesting perspectives regarding the actual strength of existing services and the potential protection improvements than our approach can provide.

6.4. A methodological framework to promote the use of renewable energy

Participants: Alexandre Rio, Yoann Maurel [contact].

This work is in line with projects aimed at optimizing the use of renewable energies. It is carried out in collaboration with OKWind. This compagny designs and supplies its customers with renewable energy generators such as vertical axis wind turbines and solar trackers. OKWind promotes a micro-grid infrastructure development.

Our application domains are those of agriculture and industry in which it is possible to identify and influence consuming processes. We mainly consider local generation for self-consumption purposes (microgrid) as it limits infrastructure costs, minimizes line losses, reduces the need of the Grid and hopefully reduces the electricity bill.

Renewable energies currently benefit from numerous subsidies to promote their use so as to reduce greenhouse gas emissions. Nevertheless, it seems worth considering the cost-effectiveness of these solutions without these incentives, as they are highly dependent on political will and can be questioned. The reduction in manufacturing costs, particularly in solar energy, suggests that these solutions can eventually compete with traditional sources if they are properly used.

Competitive low-carbon energy is hampered by the stochastic nature of these sources. During peak periods, the electricity produced is competitive, but too often, the scheduled consumption is not aligned with production. In practice, process planning was and is still driven by the electricity price from the grid. On average, the profitability of the installations is therefore not certain. In this context, using battery to shift the load looks appealing but is, as of today, far from being economically viable if not done properly.

Consequently, the achievement of a profitable self-production site is, in practice, a question of trade-off that involves several factors: the scaling of energy sources, the sizing of batteries used, the desired autonomy level, the ecological concerns, and the organization of demand. This trade-off analysis is very challenging: to be carried out effectively and comprehensively, it must be supported by tools that help the stakeholders. While much work has been done in the literature on the impacts of different factors, there are few approaches that offer a comprehensive model.

Our objective is to provide a methodological framework to embrace the diversity of knowledge, of production and consumption tools, of farm activities and of prediction algorithms. This should enable an expert to conduct a trade-off analysis and decide on the best option for each individual site under consideration.

In the first two years of this PhD thesis, we argued that model-driven engineering is suited for the development of such a model and we presented some preliminary implementation. In 2019, we were able to test our approaches in the field and continue to expand the model to account for a wider range of resources. This was published in [5].

6.5. Introducing Data Quality to the Internet of Things

Participants: Jean-Marie Bonnin, Frédéric Weis [contact].

The Internet of Things (IoT) connects various distributed heterogeneous devices. Such Things sense and actuate their physical environment. The IoT pervades more and more into industrial environments forming the so-called Industrial IoT (IIoT). Especially in industrial environments such as smart factories, the quality of data that IoT devices provide is highly relevant. However, current frameworks for managing the IoT and exchanging data do not provide data quality (DQ) metrics. Pervasive applications deployed in the factory need to know how data are "good" for use. However, the DQ requirements differ from a process to another. Actually, specifying/expressing DQ requirements is a subjective task, depending on the specific needs of each targeted application. As an example this could mean how accurate a location of an object that is provided by an IoT system differs from the actual physical position of the object. A Data Quality of 100% could mean that the value represents the actual position. A Data Quality of 0% could mean that the object is not at the reported position. In this example, the value 0% or 100% can be given by a specific software module that is able to filter raw data sent to the IoT system and to deliver the appropriate metric for Dev apps. Building ad hoc solutions for DQ management is perfectly acceptable. But the challenge of writing and deploying applications for the Internet of Things remains often understated. We believe that new approaches are needed, for thinking DQ management in the context of extremely dynamic systems that is the characteristic of the IoT.

In 2019, we introduced DQ to the IoT by (1) representing data quality parameters as metadata to each stored and exchanged IoT data item and (2) providing a toolbox that helps developers to assess the data quality of their processed data using the previously introduced data quality metadata. We followed an inductive approach. Therefore, we set up a pilot to gain first-hand experience with DQ, and to test our developed tools. Our pilot focuses on multi-source data inconsistency. Our setting consists of multiple industrial robots that cowork within a factory. The robots on the line follow a fixed path while the other two robots can freely move. For our implementation we use a data-centric IoT middleware, the Virtual State Layer (VSL). It provides many desired properties such as security and dynamic coupling of services at runtime. Most important it has a strong semantic model for representing data that allows adding new metadata for data quality easily. In our pilot the decrease of the DQ is caused by a low periodicity of location reports. We implemented a DQ service that infers the DQ being located in the service chain. The coordination service queries our DQ enriching service. The DQ enrichment service models the behavior of a robot and infers the resulting DQ depending on the time between the location report and the coordination service's query. Our goal was not only to report the DQ to the consuming service but also to offer tools (microservices) to mitigate from bad DQ. To enable a mitigation from the decreasing DQ, we started the sensors at a random time. This results in the same precision decrease periodicity but in shifted reporting times. The shift enables increasing the DQ by using sensor fusion and data filtering.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

Project: SIMHet

Partner: YoGoKo

Coordinator: JM. Bonnin

Starting: Nov 2015 - Ending: April 2020

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.

7.2. Bilateral Grants with Industry

OKWIND

Coordinator: Y. Maurel

Starting: April 2017 - Ending: April 2020

Abstract: OKWind [0] is a company specialized in local production of renewable energy. This project, with Inria DiverSE and EASE teams, 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, so as to favor the consumption of locally produced electricity. The system must be able to model clients' activities. It must also trigger actions (local consumption vs. local storage). The final goal is to use "locally produced" energy in a smarter way and to tend towards a self-consumption optimum. This contract funds Alexandre Rio's PhD grant.

[0] http://www.okwind.fr/

Orange Labs

Coordinator: JM. Bonnin

Starting: Jan 2016 - Ending: Jan 2019

Abstract: The objective of this thesis is to propose a new management architecture for optimizing the upstream bandwidth allocation in PON while acting only on manageable parameters to allow the involvement of self-decision elements into the network. To achieve this, classification techniques based on machine learning approaches are used to analyze the behavior of PON users and specify their upstream data transmission tendency. A dynamic adjustment of some SLA parameters is then performed to maximize the overall customers' satisfaction with the network. This contract funds Nejm Frigui's PhD grant, co-supersized with Tayeb Lemlouma (IRISA OCIF team).

8. Partnerships and Cooperations

8.1. Regional Initiatives

Chantier 3.0

Coordinator: JM. Bonnin

Starting: Jan 2019; Ending : Dec 2021

Partners: Agemos, YoGoKo, IMT Atlantique

Abstract: Co-founded by "Région Bretagne" Chantier 3.0 is a "PME Project" aiming at increasing safety of workers in construction sites and road works. In these scenarios, vehicles represent a danger for the workers. Knowing the position of the vehicles and workers, it is possible to alert workers who are located in a safety perimeter around the vehicles. The project addresses the challenges of 1) precise localisation with low or medium cost wearable devices and 2) of dynamically setting up a reliable communication network in harsh environments mixing indoor and outdoor conditions. The key technologies used to solve theses issues include: fusion of localisation data (GPS, acceleration integration, location anchors, angle of arrival and time of flight of radio signals), opportunistic short range broadcast communications, ITS communication protocols and system integration. EASE brings it expertise in all of theses domains in order to enhance the reliability of the system, to make it affordable and to pave the way for its standardisation.

8.2. National Initiatives

SCOOP@F part 2

Coordinator: JM. Bonnin

Starting: Jan 2016; Ending: Dec 2019

Partners: MEDE, Renault, PSA, IMT Atlantique

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 and SCOOP@F Part 2 from 2016 to 2019. 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). The second phase of SCOOP will end up in 2019. As a partner of the InDiD consortium, we proposed a follow up for this project to the EC for the period 2020-2023.

InDiD

Coordinator: JM. Bonnin

Starting: mid 2019; Ending: Dec 2023

Partners: 20+ French partners including cities (Paris, Grenoble...), road operators, transport operators, academics (incl. IMT Atlantique) and industrials

Abstract: InDiD is one of 13 French projects out of 148 European projects selected by the European Commission within the framework of the last Connecting Europe Facility (CEF) call for proposals. The project benefits from a co-funding rate of 50% on behalf of the European Union. It follows the Smart Cooperative Transport Systems projects SCOOP@F, C-ROADS France and InterCor. The project aims at expanding the coverage of use cases deployed in previous projects (emergency braking, accident, work...) and develop new use cases dealing with urban area, but also use cases of increased perception for autonomous vehicle. In addition, it deals with high definition digital mapping of the infrastructure. Connectivity along with mapping shape the digital infrastructure of tomorrow, an essential addition to the physical infrastructure. InDiD aims at continuing the deployment of Cooperatives Intelligent Transport Systems on new road experimentation sites in order to expand the services coverage offered by the infrastructure. Pilot sites are located on 4 main French geographic areas, on the Mediterranean side, in the south-west area, at the centre and in the north of France.

TAGRI

Coordinator: P. Couderc

Starting: Nov 2019; Ending: Nov 2020

Abstract: Tagri is a 12 months innovation action supported by a CominLabs grant, started on 2019-11-01 and ending on October 2020. It follows up the previous Pervasive_RFID project, a joint Inria - IETR collaboration. Tagri aims at developing an operational UHF RFID solution for agricultural applications where tags are used as a pervasive storage to track important data related to the production. Tagri is using the RFID research facility from Pervasive_RFID project to study the behavior and performance level of UHF RFID in the context of agricultural applications, which is new as the the standard RFID technology used in farming is LF based: historically, LF was selected because it was reliable for bio-tags attached to animal, and the driver application for RFID in smart farming was breeding. A new research engineer, Alexis Girard, has integrated the team in November 2019 on this project.

8.3. International Research Visitors

8.3.1. Informal International Partners

Three years ago we initiated a collaboration with Valerie Gay and Christopher Lawrence (UTS / Australia) on adapting smart spaces for eHealth applications. We continued the collaboration and Jean-Marie Bonnin visited UTS last August. He participated in the definition of research IoT infrastructure for a new maternity clinic dedicated to aboriginal community. The goal was to design an efficient research infrastructure to study how pervasive technologies could be used to adapt the environment to the people. To prepare this visit, Christopher Lawrence came in France and visit the team in March 2019.

8.3.2. Visits of International Scientists

Christopher Lawrence, Associate Professor, University of Technology Sydney, visited the team in March/April 2019.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Organisation

9.1.1.1. Member of the Organizing Committees

The E4SE team has been involved from the beginning with the City of Rennes in the design of the inOut event and its experts comitee. inOut is an innovative initiative at the service of new mobility solutions. After holding a first event in March 2018, inOut returned to Rennes in 2019 from 28 to 31 March and will happen again next years. The two-day [IN]door business event, followed by an [OUT]door event open to the general public for the entire weekend, gathers together mobility businesses and users to debate, test, and explore new mobility solutions. At the crossroads of digital technologies and mobilities, inOut seeks to bring out the best these two worlds have to offer by providing a space of discussion and debate, where topics and concepts are unpicked and decoded, and different ideas converge. In addition to the expert-focused [IN] part of the event which targets mobility professionals from all over the world, there is also an [OUT] section that invites the general public to take a weekend to explore and discover this topic of new mobilities. Field tests and experimentations are also conducted as part of inOut to support start-ups, businesses and manufacturers and to encourage them to explore what the future holds.

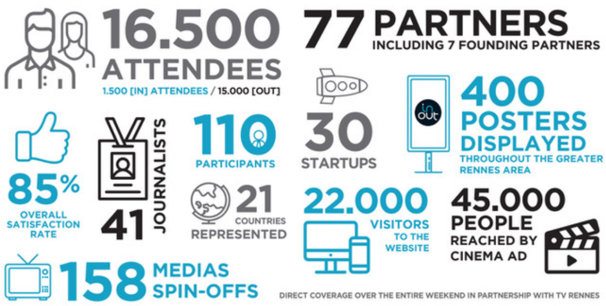


Figure 4. inOut 2019 in figures

inOut is not only an event and we built in with experimentation at the heart of the innovation approach. It has a life beyond the event. It offers an ecosystem of new mobility solutions throughout the year. Experiments, demonstrations, hackathons, and academic challenges get students, researchers, start-ups, and users involved. The call for experiments encourages entrepreneurs and academics to create and test innovations all along the year. An inOut experimentation is conducted in an open-air laboratory: the city of Rennes places its streets and facilities at the disposal of start-ups and companies so that they can test the digital technologies of the city of tomorrow. It aims to provide a response to the need for innovation in communities, to upgrade and develop business sectors, to improve the appeal of the region (companies and universities) for companies and research projects.

9.1.2. Scientific Events: Selection

9.1.2.1. Member of the Conference Program Committees

- PC member of VTC 2019, ICIN 2019, ICON 2019, VEHITS 2020, ICCVE 2019, 3ICT 2019, JM. Bonnin
- PC member of eHPWAS 2019, AdHoc Now 2020, F. Weis

9.1.3. Journal

9.1.3.1. Reviewer - Reviewing Activities

- IEEE Communication Magazine, JM. Bonnin and C. Couturier
- IEEE Access, Wireless Communication and Mobile Computing, The Computer Journal, Transactions on Mobile Computing, Vehicular communications, Journal of sensors, JM. Bonnin
- IEEE Computer, F. Weis

9.1.4. Invited Talks

- Presentation at Security Days @ IRISA, JM Bonnin
- Round table at InOut 2019, JM Bonnin
- Presentation of the EASE activity around Mobile Data collection in Smart City at South West Aboriginal Medical Service, JM Bonnin

9.1.5. Scientific Expertise

- Head of the scientific committee of InOut 2019, JM. Bonnin
- Member of the scientific council of the GIS ITS, JM. Bonnin
- Member of the scientific council the Id4Car cluster, JM. Bonnin
- Scientific advisor of the YoGoKo startup, JM. Bonnin
- Scientific advisor for an INRA team (Unité Expérimentale Ecologie et Ecotoxicologie Aquatiques) on fish tracking issues using LF RFID, P. Couderc
- Co-head of "the pole Digital Society of the MSHB" (Maison des Sciences de l'Homme de Bretagne), JM. Bonnin
- Project evaluation for ANR, Belgium, Id4Car, Région Pays de la Loire, Région Grand Est, JM. Bonnin
- Expert for CSV board of "Pôle Images et Réseaux", projects reviewing and selection, strategic roadmap definition, P. Couderc

9.1.6. Research Administration

- Head of the Networks, Telecommunication and service department at IRISA, JM. Bonnin until Nov 2019
- Member for the IRISA Laboratory Council, JM. Bonnin
- Member of the scientific council of ECAM (Engineering school), 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 2: Wireless LANs, F. Weis, 30 hours, M2, IMT Atlantique
- Master 2: Pervasive computing and IoT system architectures, 4 hours, P. Couderc, Univ. Rennes 1
- Master 1: Network programming (lectures, tutorial, labs), 78 hours, Y. Maurel, Univ. Rennes 1
- 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
- Master 2 : Mobility management in the Internet, JM. Bonnin, IMT Atlantique
- Master 2 Smart Mobility : Communications for ITS, JM. Bonnin, IMT Atlantique
- Master 2 IoT: Smart City, JM. Bonnin, ENSI Tunis

- Continuous training: Communications for Autonomous and Cooperative vehicle, Communications for ITS, JM. Bonnin, IMT Atlantique
- Master 1: Network programming (lab and project), 72 hours, C. Couturier, IMT Atlantique
- Master 1: IP networks (lectures, tutorial), 24 hours, C. Couturier, IMT Atlantique
- Master 2: Enterprise Network Design, 12 hours, C. Couturier, IMT Atlantique
- Master 2: Bid to Call for Tender, 12 hours, C. Couturier, IMT Atlantique
- Master 2: Supervision of the Networking and Telecom curriculum of FIP3A, C. Couturier, IMT Atlantique
- Master 2: Tools for cloud computing , 18 hours, C. Couturier, IMT Atlantique

9.2.2. Supervision

PhD in progress: Indra Ngurah, Car-based Data Collection for Low Energy Devices (Car-based DC4LED), May 2016, Jean-Marie Bonnin

PhD in progress: Alexandre Rio, "Modélisation des activités de site consommateur d'énergie", Octobre 2016, Olivier Barais (UR1) and Yoann Maurel

PhD in progress: Rodrigo Silva, "Mécanisme de décision multi-critères pour le placement de flux en environnement hétérogène et changeant", Nov. 2015, Jean-Marie Bonnin

PhD in progress: Nejm Frigui, "Maintenance autonomique du réseau programmable d'accès optiques de très haut débit", Jan. 2016, Jean-Marie Bonnin

9.2.3. Juries

PhD: Mohammad Irfan Khan, "Multi-Service Resource Orchestration for Vehicular Safety Communication", Jean-Marie Bonnin, referee Dec 2019

10. Bibliography

Publications of the year

International Conferences with Proceedings

[1] P. COUDERC, Y. MAUREL.*Location corroboration using passive observations of IEEE 802.11 Access Points*, in "CCNC 2019 - 16th IEEE Consumer Communications & Networking Conference", Las Vegas, United States, IEEE, January 2019, p. 1-7 [*DOI :* 10.1109/CCNC.2019.8651873], https://hal.inria.fr/hal-01954122

[2] N. I. ER, K. D. SINGH, J.-M. BONNIN.*DC4LED: A Hierarchical VDTN Routing for Data Collection in Smart Cities*, in "CCNC 2019 - 16th IEEE Consumer Communications and Networking Conference", Las Vegas, United States, Proceeding of The 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC 2019), IEEE, April 2019, vol. 1, p. 883-886 [*DOI :* 10.1109/CCNC.2019.8651757], https://hal.archives-ouvertes.fr/hal-01894719

[3] R. SILVA, C. COUTURIER, J.-M. BONNIN, T. ERNST.*A Heuristic Decision Maker Algorithm for Opportunistic Networking in C-ITS*, in "VEHITS 2019 - 5th International Conference on Vehicle Technology and Intelligent Transport Systems", Heraklion, Greece, SCITEPRESS - Science and Technology Publications, May 2019, p. 578-585 [*DOI :* 10.5220/0007799005780585], https://hal-imt-atlantique.archives-ouvertes.fr/hal-02177505

Conferences without Proceedings

[4] N. I. ER, K. D. SINGH, J.-M. BONNIN.*CityView: A Vehicle Based Image and Data Collection for Smart Cities*, in "WiMob 2019 - 15th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications", Barcelona, Spain, October 2019, p. 1-7 [*DOI :* 10.1109/WIMOB.2019.8923547], https://hal.archives-ouvertes.fr/hal-02318697

[5] A. RIO, Y. MAUREL, Y. BUGNI, O. BARAIS.*Benefits of Energy Management Systems on local energy efficiency, an agricultural case study*, in "SmartGridComm 2019 - IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids", Beijing, China, IEEE, October 2019, p. 1-7, https://hal.archives-ouvertes.fr/hal-02315327

Scientific Books (or Scientific Book chapters)

[6] R. SILVA, C. COUTURIER, T. ERNST, J.-M. BONNIN.*Proactive Decision Making for ITS Communication*, in "Global Advancements in Connected and Intelligent Mobility: Emerging Research and Opportunities", IGI Global, July 2019, p. 197-226 [*DOI :* 10.4018/978-1-5225-9019-4.CH006], https://hal-imt-atlantique.archives-ouvertes.fr/hal-02173428

Project-Team EMPENN

EMPENN

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

Project-Team EMPENN

Creation of the Project-Team: 2019 January 01

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. Team, Visitors, External Collaborators

Research Scientists

Christian Barillot [Team leader, CNRS, Senior Researcher, HDR]
Olivier Commowick [Inria, Researcher, HDR]
Emmanuel Caruyer [CNRS, Researcher]
Julie Coloigner [CNRS, Researcher]
Camille Maumet [Inria, Researcher]

Faculty Members

Pierre Maurel [Team leader, Univ de Rennes I, Associate Professor]
Isabelle Bonan [Univ de Rennes I, Professor, HDR]
Gilles Edan [Univ de Rennes I, Professor, HDR]
Jean-Christophe Ferré [Univ de Rennes I, Professor, HDR]
Jean-Yves Gauvrit [Univ de Rennes I, Professor, HDR]

Technical Staff

Élise Bannier [Centre hospitalier régional et universitaire de Rennes, Engineer]
Benoit Combès [Inria, Engineer]
Aurelien Cornet [INSERM, Engineer, from May 2019]
Isabelle Corouge [Univ de Rennes I, Engineer]
Arthur Masson [INSERM, Engineer, from Jul 2019]

PhD Students

Mathis Fleury [Inria]
Pierre-Yves Jonin [Centre hospitalier régional et universitaire de Rennes, until Sep 2019]
Antoine Legouhy [CNRS]
Giovanna Orru [Univ de Rennes I, from Nov 2019]
Maia Proisy [Centre hospitalier régional et universitaire de Rennes, until Apr 2019]
Xavier Rolland [CNRS]
Raphaël Truffet [Univ de Rennes I]
Corentin Vallée [Univ de Rennes I]
Stephanie Leplaideur [Centre hospitalier régional et universitaire de Rennes]
Cédric Meurée [Siemens CIFRE, until March 2019]
Haykel Snoussi [Inria, until May 2019]

Post-Doctoral Fellows

Hector Garcia [CNRS, from Oct 2019]
Claire Cury [Inria, Engineer]
Quentin Duché [Inria, Engineer]
Francesca Galassi [Inria, Engineer, until Aug 2019]
Giulia Lioi [Inria, Engineer]

Visiting Scientists

Gabriel Besson [Mobility grant of Wallonie-Bruxelles International, from Jul 2019 until Sep 2019]
Gregory Kiar [McGill University, from Apr 2019 until Jun 2019]

2. Overall Objectives

2.1. Overall Objectives

Empenn (means “Brain” in Breton language) ERL U1228 research team is jointly affiliated with Inria, Inserm (National Institute of Health and Scientific Research), CNRS (INS2I institute), and University of Rennes I. It is a team of IRISA/UMR CNRS 6074. Empenn is based in Rennes, at both the medical and science campuses. The team follows the “VisAGeS” one that was created for 12 years in 2006 by Inria, As for “VisAGeS”, Empenn hosts the accreditation number U1228 renewed by Inserm in 2017, after a competitive evaluation conducted by both HCERES and Inserm.

Through this unique partnership, the ambition of Empenn is to establish a multidisciplinary team bringing together researchers in information sciences and medicine. Our medium- and long-term objective is to introduce our basic research to clinical practice, while maintaining the excellence of our methodological research.

Our goal is to foster research in medical imaging, neuroinformatics and population cohorts. In particular, the Empenn team targets the detection and development of imaging biomarkers for brain diseases and focus its efforts on translating this research to clinics and clinical neurosciences at large.

In particular, the objective of Empenn is to propose new statistical and computing methods, and to measure and model brain morphological, structural and functional states in order to better diagnose, monitor and deliver treatment for mental, neurological and substance use disorders. We propose combining advanced instrumental devices and new computational models to provide advanced diagnosis, therapeutic and neuro-rehabilitation solutions for some of the major disorders of the developing and aging brain.

Generic and challenging research topics in this broad domain include finding new ways to compare models and data, assist decisions and interpretation, and develop feedback from experiments. These activities are performed in close collaboration with the Neurinfo in vivo imaging platform, which is a critical environment for the experimental implementation of our research on challenging clinical research projects and the development of new clinical applications.

3. Research Program

3.1. Scientific Foundations

The scientific foundations of our team concern the design and development of new computational solutions for biological images, signals and measurements. Our objective is to develop a better understanding of the normal and pathological brain, at different scales.

This includes imaging brain pathologies in order to better understand pathological behavior from the organ level to the cellular level, and even to the molecular level (using molecule (e.g. through PET-MR imaging), as well as modeling with specific ligands/nanocarriers), and the modelling of normal and pathological large groups of individuals (cohorts) from image descriptors. It also includes the challenge of the discovery of episodic findings (i.e. rare events in large volumes of images and data), data mining and knowledge discovery from image descriptors, the validation and certification of new drugs from imaging features, and, more generally, the integration of neuroimaging into neuroinformatics through the promotion and support of virtual organizations of biomedical actors by means of e-health technologies.

As shown in Fig. 1, the research activities of the Empenn team closely link observations and models through the integration of clinical and multiscale data, and phenotypes (cellular, and later molecular, with structural or connectivity patterns in the first stage). Our ambition is to build personalized models of central nervous system organs and pathologies, and to compare these models with clinical research studies in order to establish a quantitative diagnosis, prevent the progression of diseases and provide new digital recovery strategies, while combining all these research areas with clinical validation. This approach is developed within a translational framework, where the data integration process to build the models is informed by specific clinical studies, and where the models are assessed regarding prospective clinical trials for diagnosis and therapy planning. All of these research activities will be conducted in close collaboration with the Neurinfo platform, which benefited in 2018 from a new high-end 3T MRI system dedicated to research (3T Prisma™ system from Siemens), and through the development in the coming years of multimodal hybrid imaging (from the currently available EEG-MRI, to EEG-NIRS and PET-MRI in the future).

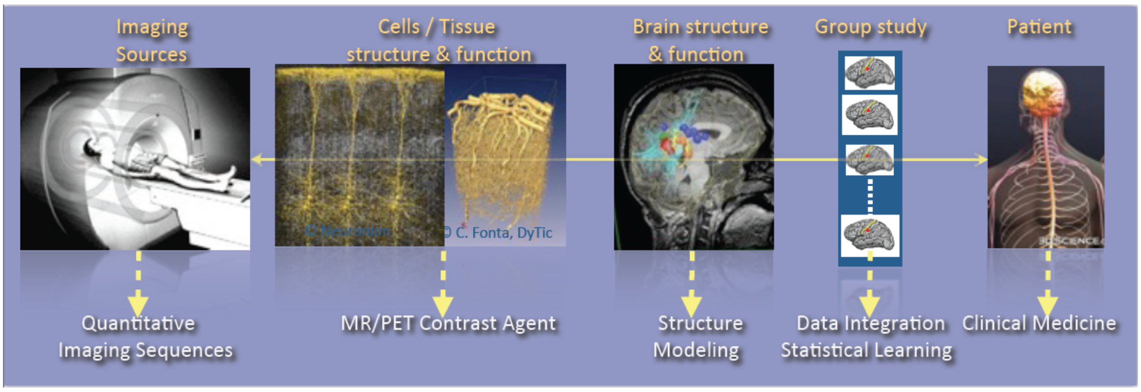


Figure 1. The major overall scientific foundation of the team concerns the integration of data from the Imaging source to the patient at different scales: from the cellular or molecular level describing the structure and function, to the functional and structural level of brain structures and regions, to the population level for the modelling of group patterns and the learning of group or individual imaging markers.

In this context, some of our major developments and newly arising issues and challenges will include:

- The generation of new descriptors to study brain structure and function (e.g. the combination of variations in brain perfusion with and without a contrast agent; changes in brain structure in relation to normal, pathological, functional or connectivity patterns; or the modeling of brain state during cognitive stimulation using neurofeedback).
- The integration of additional spatiotemporal and hybrid imaging sequences covering a larger range of observations, from the molecular level to the organ level, via the cellular level (arterial spin labeling, diffusion MRI, MR relaxometry, MR fingerprinting, MR cell labeling imaging, MR-PET molecular imaging, EEG-MRI functional imaging, EEG-NIRS-MRI, etc.).
- The creation of computational models through the data fusion of molecular, cellular (i.e. through dedicated ligands or nanocarriers), structural and functional image descriptors from group studies of normal and/or pathological subjects.
- The evaluation of these models in relation to acute pathologies, especially for the study of degenerative, psychiatric, traumatic or developmental brain diseases (primarily multiple sclerosis, stroke, traumatic brain injury (TBI) and depression, but applicable with a potential additional impact to epilepsy, Parkinson's disease, dementia, Posttraumatic stress disorder, etc.) within a translational framework.

In terms of new major methodological challenges, we will address the development of models and algorithms to reconstruct, analyze and transform the images, and to manage the mass of data to store, distribute and "semanticize" (i.e. provide a logical division of the model's components according to their meaning). As such, we expect to make methodological contributions in the fields of model inference; statistical analysis and modeling; the application of sparse representation (compressed sensing and dictionary learning) and machine learning (supervised/unsupervised classification and discrete model learning); data fusion (multimodal integration, registration, patch analysis, etc.); high-dimensional optimization; data integration; and brain-computer interfaces. As a team at the frontier between the digital sciences and clinical research in neuroscience, we do not claim to provide theoretical breakthroughs in these domains but rather to provide significant advances in using these algorithms through to the advanced applications we intend to address. In addition, we believe that by providing these significant advances using this set of algorithms, we will also contribute to exhibiting new theoretical problems that will fuel the domains of theoretical computer sciences and applied mathematics.

In summary, we expect to address the following major challenges:

- Developing new information processing methods able to detect imaging biomarkers in the context of mental, neurological, and substance use disorders.
- Providing new computational solutions for our target applications, allowing a more appropriate representation of data for image analysis and the detection of biomarkers specific to a form or grade of pathology, or specific to a population of subjects.
- Providing, for our target applications, new patient-adapted connectivity atlases for the study and characterization of diseases from quantitative MRI.
- Providing, for our target applications, new analytical models of dynamic regional perfusion, and deriving indices of dynamic brain local perfusion from normal and pathological populations.
- Investigating whether the theragnostics paradigm of rehabilitation from hybrid neurofeedback can be effective in some behavioral and disability pathologies.

These major advances will be primarily developed and validated in the context of several priority applications in which we expect to play a leading role: multiple sclerosis, stroke rehabilitation, and the study and treatment of depression.

4. Application Domains

4.1. Population imaging

One major objective of neuroimaging researchers and clinicians is to be able to stratify brain imaging data in order to derive new and more specific population models. In practice this requires to set up large-scale experiments that, due to the lack of resources and capabilities to recruit locally subjects who meet specific inclusion criteria, motivates the need for sharing the load.

But, building and using multi-site large-scale resources poses specific challenges to deal with the huge quantity of data produced and their diversity. Empenn will focus on two challenges in particular:

- Provide computational environments for the computation and use of imaging biomarkers in the targeted brain diseases, a solution to be used by radiologists and neurologists/psychiatrists for the clinical follow-up of a large patient population.
- Modeling analytic variability of image processing pipelines to better understand and predict the behaviour of imaging biomarker detection solutions and improve reproducibility and productivity in clinical neuroimaging research.

4.2. Detection and learning

We intend to make significant contributions with major impacts in learning coupling models between functional recordings during neurofeedback procedures. These advances will provide a breakthrough in brain-computer interfaces for rehabilitation protocols. Our aim is to:

- Provide a computational environment that combines data-driven (machine learning) and Bayesian solutions to improve the detection of abnormal patterns in images through decision or evidence theory data fusion strategies. The major initial application will be for multiple sclerosis. Over the longer term, we also expect to adapt these methods to address a wider range of neurological diseases (epilepsy, stroke, tumors, etc.) in neonate and adult brains.
- Develop solutions for combining brain state measurements from multimodal sensors or sequences (e.g. fMRI, ASL, EEG, NIRS, etc.) with applications in the spatiotemporal reconstruction of brain activity from MRI-EEG or the combined detection of the endogenous hemodynamic and resting state network of the brain from ASL and NIRS. Over the longer term, the advent of new hybrid brain imaging sensors (e.g. PET-MRI) will require these methods to be extended to a larger spectrum of information combining structural, morphological, metabolic, electrophysiological and cellular/molecular information (e.g. through the use of specific ligands/nanocarriers).

4.3. Quantitative imaging

The Empenn research group focuses on the development of several quantitative techniques in magnetic resonance imaging of the brain. These methods allow for a characterization of both the function and the structure of the brain with high precision. Arterial spin labelling (ASL) is a contrast agent-free imaging technique which labels arterial blood water as an endogenous tracer for perfusion and can measure resting-state cerebral blood flow. We are interested in estimating multiparametric hemodynamics using ASL, such as combined cerebral blood flow and arterial transit times, and derive statistical descriptors to represent significant differences between groups. In addition to quantitative perfusion parameters, our contributions on tissue compartment imaging aim at delineating neural circuits and characterize their microstructure properties, using both diffusion MRI and relaxometry. In diffusion MRI, arbitrary gradient waveforms were shown to exhibit higher sensitivity to microstructure parameters than standard pulsed gradients. We work on the optimization of sampling protocols in this domain, with the objective to propose sequences compatible with in vivo acquisition. Complementary to diffusion MRI, we develop methods for the reconstruction of myelin-bound, extra-axonal and cerebrospinal fluid water using multi-compartment modelling of the T2-relaxometry signal. We combine these techniques with tractography to identify trajectories of pathologies associated to the evolution of these microstructural parameters along specific fiber bundles in the brain white matter.

4.4. Behavior

Advances in the field of in vivo imaging offer new opportunities for addressing the management of resistant affective disorders and their consequences (suicide risk and socio-professional impact), and the management of spatial cognition disorders after stroke and their consequences (postural perturbations and the loss of autonomy). Our objective, and the main challenge in this context, will be to introduce medical image computing methods to the multidisciplinary field of behavioral disorders (cognitive disorders, particularly spatial and postural control disorders or anterograde memory impairment, mood disorders, notably resistant depression, schizophrenic disorders, pervasive developmental disorders, attention disorders, etc.) in order to gain a better understanding of the pathology and devise innovative therapeutic approaches.

We also expect to become a major player in the future and make important contributions with significant impacts, primarily in drug-resistant depression in young and old populations. In particular, we expect to provide new image-related metrics combining perfusion, metabolism and microstructural information regarding the brain in order to better characterize pathologies, provide prospective evolution values and potentially provide new brain stimulation targets that could be used in neurofeedback rehabilitation protocols or other types of brain stimulation procedure.

We aim to provide new imaging markers of mental diseases, especially in the context of mood disorders. The new biomarkers will be derived from the metabolic (ASL and later ASL+PET) point of view as well as from the microstructural point of view (multicompartment diffusion MRI and relaxometry). Similarly, we expect to exhibit imaging biomarker regularities combining metabolic and structural information. Over the longer term, we expect these biomarkers to be the target of neurofeedback rehabilitation procedures. Also, over the longer term, we expect to supplement the MRI markers with molecular marker ones coming from new PET tracers, especially those associated with serotonin intake, at one time point or during a rehabilitation protocol under hybrid PET-EEG-MRI neurofeedback procedures.

4.5. Neuroinflammation

Some of the major ongoing research issues in the neuroimaging of neuro-inflammatory diseases concern the definition of new biomarkers for tracking the development of the pathology using high- dimensional data (e.g. nD+t MRI). This includes the use of white matter-specific imaging, such as magnetization transfer MRI, relaxometry and diffusion-weighted imaging (DW-MRI). Our objective is (1) to develop information-processing tools to tag the spatiotemporal evolutions of MS patterns at the brain parenchyma and spinal cord levels from their different signatures (inflammatory cells visible with USPIO or Gd contrast agents on MRI, persistent black holes, eloquent regional atrophy and microstructure signatures); and (2) to test these new tools on new imaging cohorts. In this respect, we for instance conduct studies on brain and spinal cord imaging, continuing on from the PHRC multicentric EMISEP project (PI: G. Edan), as it is very likely that lesions in the spine will directly affect the ambulatory ability of the patient (and thereby the clinical scores). In order to extend this experiment to a larger MS population, based on our expertise from the OFSEP cohort, we also plan to improve the MS therapeutic decision process through the MUSIC project (Multiple Sclerosis Imaging Check out, a public/private project). Our goal is to develop and assess a standardized monitoring tool that provides a robust, long-term computerized MRI follow-up that will become the gold standard in clinical practice for therapeutic decisions in MS treatment. As part of this project, Empenn will share its expertise in data management systems (Shanoir and FLI-IAM) and automatic processing tools (through the medInria and Anima software repositories) to extract quantitative indices from the images.

4.6. Recovery

Mental and neurological disorders are the leading cause of years lived with a disability. Treatment-resistant depression affects approximately 2% of the European population. Meanwhile, in the case of brain disorders, almost 1.5 million Europeans (15 million people worldwide) suffer a stroke event each year. Current recovery methods for brain disorders and traumatic brain injuries remain limited, preventing many from achieving full recuperation. We propose addressing the issue of brain recovery by introducing new advances from recent breakthroughs in computational medical imaging, data processing and human-machine interfaces, and demonstrating how these new concepts can be used, in particular for the treatment of stroke and major depressive disorders.

We ambition to combine advanced instrumental devices (Hybrid EEG, NIRS and MRI platforms), with new hybrid brain computer interface paradigms and new computational models to provide neurofeedback-based therapeutic and neuro-rehabilitation paradigms in some of the major mental and neurological disorders of the developmental and the aging brain.

Neurofeedback involves using a brain-computer interface that provides an individual with real-time biofeedback about his or her brain activity in the form of sensory feedback. It enables individuals to learn to better control their brain activity, which can be measured in real time using various non-invasive sensors as described above. Although EEG is currently the only modality used by clinical practitioners in that context, it lacks specificity due to its low spatial resolution. Dynamic research into fMRI-neurofeedback has held promise for treating depression, chronic pain and stroke, since it offers the prospect of real-time imagery of the activity in deep brain structures with high spatial resolution. However, the low temporal resolution and high cost of fMRI-Neurofeedback has hampered the development of many applications. We believe that the future belongs to hybrid responses that combine multimodal sensors and intend to demonstrate this in the Empenn project.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. New NIRS system at the Neurinfo platform

An MRI and EEG-compatible functional near-infrared spectroscopy (fNIRS) system was installed at the Neurinfo platform in September 2019.

5.1.2. Sciences en Cour[t]s

This event is a festival of short films, which offers doctoral students the opportunity to make short films about their thesis work. Raphael Truffet, Antoine Legouhy and Xavier Rolland won the high school award in science en Cour[t]s event https://www.youtube.com/watch?v=IKgqv-iCwak.

5.1.3. Second neuroscience hackathon in Rennes

We organized the second edition of hackathon in the Empenn team, November 14-15 as part of the international event Brainhack Global 2019.

6. New Software and Platforms

6.1. Anima

KEYWORDS: Filtering - Medical imaging - Diffusion imaging - Registration - 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, 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 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, Pierre Maurel, Quentin Duché and Julie Coloigner
- Contact: Isabelle Corouge
- URL: https://team.inria.fr/visages/software/

6.3. MedInria

KEYWORDS: Visualization - DWI - Health - Segmentation - Medical imaging

SCIENTIFIC DESCRIPTION: MedInria 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 Empenn 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: https://med.inria.fr

6.4. QtShanoir

KEYWORDS: Qt - Nifti - Medical imaging - Plug-in - DICOM - Health - C++ - Soap - Webservices - Shanoir

SCIENTIFIC DESCRIPTION: QtShanoir is based on Qt/C++ library. It interacts with the Shanoir server using SOAP web services. 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 : search, 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 share, archive, search and visualize neuroimaging data.

It provides a user-friendly secure web access and offers an intuitive workflow to facilitate the collecting and retrieving of neuroimaging data from multiple sources and a wizzard to make the completion of metadata easy. Shanoir comes along many features such as anonymization of data, support for multi-centric clinical studies on subjects or group of subjects.

Shanoir offers an ontology-based data organization (OntoNeuroLOG). Among other things, this facilitates the reuse of data and metadata, the integration of processed data and provides traceability through an evolutionary approach. Shanoir allows researchers, clinicians, PhD students and engineers to undertake quality research projects with an emphasis on remote collaboration. As a secured J2EE web application, it therefore allows safely storing and archiving, with no more requirements than a computer with an internet connection!

Furthermore, Shanoir is not only a web application: it is also a complete neuroinformatics platform in which you can easily integrate your existing processing tools or develop your own ones (see ShanoirTk).

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 and 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 regard to security, the system requires authentication and user rights are tunable for each hosted study. The person responsible for the study can define which users are allowed to see, download or import data.

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.

- 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: Webservices - PACS - Medical imaging - Neuroimaging - DICOM - Health - Biology - Java - 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, Inès Fakhfakh, Justine Guillaumont, Michael Kain and Yao Yao
- Contact: Christian Barillot
- URL: http://shanoir.gforge.inria.fr

6.7. Anima medInria plugins

KEYWORDS: IRM - Medical imaging - Diffusion imaging

FUNCTIONAL DESCRIPTION: Plugins for the medInria software based on the open source software Anima developed in the Visages / Empenn team. These plugins are interfaces between anima and medinria allowing to use Anima functionalities within the clinical user interface provided by medInria. The current functionalities included in the plugins are right now: image registration, denoising, quantitative image (relaxometry), and model estimation and visualization from diffusion imaging.

- Participants: Olivier Commowick, René-Paul Debroize and Guillaume Pasquier
- Contact: Olivier Commowick

6.8. Platforms

6.8.1. The Neurinfo Platform

Empenn 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, CNRS 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 (https://www.neurinfo.org).

Concerning the Neurinfo Platform, the activity domain is a continuum between methodological and technological research built around specific clinical research projects. 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 central nervous system 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. A new 3T Siemens Prisma MRI scanner was installed at the Neuroinfo platform in February 2018. In 2014, an 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 High Performance Computing environment made of one large computing cluster and a data center that is shared and operated by the Inria center and IRISA (UMR CNRS 6074). The computation cluster (480 cores) and the data center (up to 150 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 60 sites). In 2019, an MRI and EEG-compatible fNIRS system was acquired through a co-funding from the INS2I institute of CNRS and FEDER. At the end of 2019, GIS IBISA awarded the Neurinfo platform with a complementary funding that will be dedicated to supplement the current system with additional sensors (from 8x8 optodes to 16x16 optodes).

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 following major methodological topics:

7.1.1. Diffusion imaging

7.1.1.1. Free water estimation using single-shell diffusion-weighted images

Participant: Emmanuel Caruyer.

Free-water estimation requires the fitting of a bi-compartment model, which is an ill-posed problem when using only single-shell data. Its solution requires optimization, which relies on an initialization step. We propose a novel initialization approach, called "Freewater EstimatoR using iNtErpolated iniTialization" (FERNET), which improves the estimation of free water in edematous and infiltrated peritumoral regions, using single-shell diffusion MRI data. The method has been extensively investigated on simulated data and healthy and brain tumor datasets, demonstrating its applicability on clinically acquired data. Additionally, it has been applied to data from brain tumor patients to demonstrate the improvement in tractography in the peritumoral region [57].

7.1.1.2. Multi-dimensional diffusion MRI sampling scheme: B-tensor design and accurate signal reconstruction
Participant: Emmanuel Caruyer.

B-tensor encoding enables the separation of isotropic and anisotropic tensors. However, little consideration has been given as to how to design a B-tensor encoding sampling scheme. In this work, we propose the first 4D basis for representing the diffusion signal acquired with B-tensor encoding. We study the properties of the diffusion signal in this basis to give recommendations for optimally sampling the space of axisymmetric b-tensors. We show, using simulations, that the proposed sampling scheme enables accurate reconstruction of the diffusion signal by expansion in this basis using a clinically feasible number of samples [24].

This work was done in collaboration with A. Bates, Australian National University and Al. Daducci, University of Verona.

7.1.1.3. Optimal selection of diffusion-weighting gradient waveforms using compressed sensing and dictionary learning
Participants: Raphaël Truffet, Emmanuel Caruyer, Christian Barillot.

Acquisition sequences in diffusion MRI rely on the use of time-dependent magnetic field gradients. Each gradient waveform encodes a diffusion-weighted measure; a large number of such measurements are necessary for the in vivo reconstruction of microstructure parameters. We propose here a method to select only a subset of the measurements, while being able to predict the unseen data using compressed sensing. We learn a dictionary using a training dataset generated with Monte-Carlo simulations; we then compare two different heuristics to select the measures to use for the prediction. We found that an undersampling strategy limiting the redundancy of the measures allows for a more accurate reconstruction when compared with random undersampling with similar sampling rate [49].

7.1.1.4. Geometric evaluation of distortion correction methods in diffusion MRI of the spinal cord
Participants: Haykel Snoussi, Emmanuel Caruyer, Olivier Commowick, Benoit Combès, Élise Bannier, Christian Barillot.

Acquiring and processing Diffusion MRI in spinal cord present inherent challenges. Differences in magnetic susceptibility between soft tissues, air and bones make the magnetic field non uniform in spinal cord. In this context, various procedures were proposed for correcting inhomogeneity-induced distortions; in this work, we propose novel geometric statistics to measure the alignment of the reconstructed diffusion model with the apparent centerline of the spine. In parallel of the correlation with an anatomical T2-weighted image, we show the utility of these statistics to study and evaluate the impact of distortion correction by comparing three correction methods using a pair of images acquired with reversed gradient polarity [48].

This work was done in collaboration with Anne Kerbrat, Neuropoly Montréal and Julien Cohen-Adad from NeuroPoly Lab, Institute of Biomedical Engineering, Polytechnique Montreal, Montreal, QC, Canada.

7.1.2. Arterial Spin Labeling

7.1.2.1. Acquisition duration in resting-state arterial spin labeling. How long is enough?
Participants: Corentin Vallée, Pierre Maurel, Isabelle Corouge, Christian Barillot.

Resting-state Arterial Spin Labeling (rs-ASL) is a rather confidential method compared to resting-state BOLD but it drives great prospects with respect to potential clinical applications. By enabling the study of cerebral blood flow maps, rs-ASL can lead to significant clinical subject-scaled applications as CBF is a biomarker in neuropathology. An important parameter to consider in functional imaging is the acquisition duration. Despite directly impacting practicability and functional networks representation, there is no standard for rs-ASL. Our work here focuses on strengthening the confidence in ASL as a rs-fMRI method, and on studying the influence of the acquisition duration. To this end, we acquired a long rs-ASL sequence and assessed the quality of typical functional brain networks quality over time compared to gold-standard networks. Our results show that after 14min of duration acquisition, functional networks representation can be considered as stable [58], [50].

7.1.2.2. Patch-based super-resolution of arterial spin labeling magnetic resonance images

Participants: Cédric Meurée, Pierre Maurel, Jean-Christophe Ferré, Christian Barillot.

Arterial spin labeling is a magnetic resonance perfusion imaging technique that, while providing results comparable to methods currently considered as more standard concerning the quantification of the cerebral blood flow, is subject to limitations related to its low signal-to-noise ratio and low resolution. In this work, we investigated the relevance of using a non-local patch-based super-resolution method driven by a high-resolution structural image to increase the level of details in arterial spin labeling images. This method was evaluated by comparison with other image resolution increasing techniques on a simulated dataset, on images of healthy subjects and on images of subjects diagnosed with brain tumors, who had a dynamic susceptibility contrast acquisition. The influence of an increase of ASL images resolution on partial volume effects was also investigated in this work [16].

The development of this super-resolution algorithm in the context of the PhD of Cédric Meurée founded by Siemens Healthineers conducted to a stay of one month of the PhD candidate in Erlangen, during summer 2018. This immersion into the neuro-development team allowed him to integrate the proposed solution with tools in use within this team. Part of the work also consisted in reducing the computation, a factor of 5 being achieved at the end of these four weeks.

7.1.3. Atlases

7.1.3.1. Unbiased longitudinal brain atlas creation using robust linear registration and log-Euclidean framework for diffeomorphisms

Participants: Antoine Legouhy, Olivier Commowick, Christian Barillot.

We have defined a new method to create a diffeomorphic longitudinal (4D) atlas composed of a set of 3D atlases each representing an average model at a given age. This is achieved by generalizing atlasing methods to produce atlases unbiased with respect to the initial reference up to a rigid transformation and ensuring diffeomorphic deformations thanks to the Baker-Campbell-Hausdorff formula and the log-Euclidean framework for diffeomorphisms. Subjects are additionally weighted using an asymmetric function to closely match specified target ages. Creating a longitudinal atlas also implies dealing with subjects with large brain differences that can lead to registration errors. This is overcome by a robust rigid registration based on polar decomposition. We illustrated these techniques for the creation of a 4D pediatric atlas, showing their ability to create a temporally consistent atlas [22].

This work was done in collaboration with François Rousseau, IMT Atlantique, LaTIM U1101 INSERM, Brest, France, under the ANR MAIA project.

7.1.3.2. Online atlasing using an iterative centroid

Participants: Antoine Legouhy, Olivier Commowick, Christian Barillot.

Online atlasing, i.e. incrementing an atlas with new images as they are acquired, is key when performing studies on databases very large or still being gathered. We proposed to this end a new diffeomorphic online atlasing method without having to perform again the atlasing process from scratch. New subjects are integrated following an iterative procedure gradually shifting the centroid of the images to its final position, making it computationally cheap to update regularly an atlas as new images are acquired (only needing a number of registrations equal to the number of new subjects). We evaluated this iterative centroid approach through the analysis of the sharpness and variance of the resulting atlases, and the transformations of images, comparing their deviations from a conventional method using Guimond's method. We demonstrated that the transformations divergence between the two approaches is small and stable, and that both atlases reach equivalent levels of image quality [42].

This work was done in collaboration with François Rousseau, IMT Atlantique, LaTIM U1101 INSERM, Brest, France, under the ANR MAIA project.

7.1.4. Neurofeedback

7.1.4.1. Learning bi-modal EEG-fMRI neurofeedback to improve neurofeedback in EEG only

Participants: Claire Cury, Pierre Maurel, Giulia Lioi, Christian Barillot.

In neurofeedback (NF), a new kind of data are available: electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) acquired simultaneously during bi-modal EEG-fMRI neurofeedback. These two complementary techniques have only recently been integrated in the context of NF for brain rehabilitation protocols. Bi-modal NF (NF-EEG-fMRI) combines information coming from two modalities sensitive to different aspect of brain activity, therefore providing a higher NF quality. However, the use of the MRI scanner is cumbersome and exhausting for patients. We presented, a novel methodological development, able to reduce the use of fMRI while providing to subjects NF-EEG sessions of quality comparable to the bi-modal NF sessions. We proposed an original alternative to the ill-posed problem of source reconstruction. We designed a non-linear model considering different frequency bands, electrodes and temporal delays, with a structured sparse regularisation. Results show that our model is able to significantly improve the quality of NF sessions over what EEG could provide alone. We tested our method on 17 subjects that performed three NF-EEG-fMRI sessions each [30].

7.1.4.2. Can we learn from coupling EEG-fMRI to enhance neuro-feedback in EEG only?

Participants: Claire Cury, Pierre Maurel, Christian Barillot.

Neurofeedback (NF) measures brain activation during a task, and gives back to the subject a score reflecting his/her performance that he/she tries to improve. Among noninvasive functional brain imaging modalities, the most used in NF, are electro-encephalography (EEG) and the functional magnetic resonance imaging (fMRI). EEG measures the electrical activity of the brain through channels located on the scalp, with an excellent temporal resolution (milliseconds), but has a limited spatial resolution due to the well-known ill-posed inverse problem of source reconstruction. Also NF-EEG (NF session with NF scores extracted from EEG) is not easy to control since it comes from mixtures of propagating electric potential fluctuations. Blood oxygenation level dependent (BOLD) fMRI measures neuro-vascular activity, easier to control, with an excellent spatial resolution, making NF-fMRI (NF session with NF scores extracted from BOLD-fMRI) an adequate modality for NF. However its temporal resolution is only of a few seconds, and it is a costly, exhausting for subjects and time consuming modality. Since those modalities are complementary, their combined acquisition is actively investigated, as well as the methodology to extract information from fMRI with EEG which is the easiest modality to use [Abreu et al. 2018]. Our challenge is to learn EEG activation patterns from NF-fMRI scores extracted during a NF session using coupled EEG-fMRI data (NF-EEG-fMRI) to improve NF scores when using EEG only [29].

7.1.5. Deep learning

7.1.5.1. Unsupervised domain adaptation with optimal transport in multi-site segmentation of multiple sclerosis lesions from MRI data

Participants: Antoine Ackaouy, Olivier Commowick, Christian Barillot, Francesca Galassi.

Automatic segmentation of Multiple Sclerosis (MS) lesions from Magnetic Resonance Imaging (MRI) images is essential for clinical assessment and treatment planning of MS. Recent years have seen an increasing use of Convolutional Neural Networks (CNNs) for this task. Although these methods provide accurate segmentation, their applicability in clinical settings remains limited due to a reproducibility issue across different image domains. MS images can have highly variable characteristics across patients, MRI scanners and imaging protocols. Retraining a supervised model with data from each new domain is not a feasible solution because it requires manual annotation from expert radiologists. In this work, we explored an unsupervised solution to the problem of domain shift. We presented a framework, Seg-JDOT, which adapts a deep model so that samples from a source domain and samples from a target domain sharing similar representations will be similarly segmented. We evaluated the framework on a multi-site dataset, MICCAI 2016, and showed that the adaptation towards a target site can bring remarkable improvements in a model performance over standard training [54].

This work was done in collaboration with Nicolas Courty, Obelix team, IRISA laboratory from University of Bretagne Sud.

7.1.5.2. Deep learning for multi-site MS lesions segmentation: two-step intensity standardization and generalized loss function.

Participants: Francesca Galassi, Olivier Commowick, Christian Barillot.

We presented an improved CNN framework for the segmentation of Multiple Sclerosis (MS) lesions from multi-modal MRI. It uses a two-step intensity normalization and a cascaded network with cost sensitive learning. Performance was assessed on a public multi-site data-set [35].

7.2. Research axis 2: Applications in Neuroradiology and Neurological Disorders

Our objectives is also to provide new computational solutions for our target clinical applications (radiology, neurology, psychiatry and rehabilitation...), allowing a more appropriate representation of data for image analysis and the detection of biomarkers specific to a form or grade of pathology, or specific to a population of subjects. In this section, we present our contributions in different clinical applications.

7.2.1. Rehabilitation

7.2.1.1. Efficacy of EEG-fMRI Neurofeedback for stroke rehabilitation in relation to the DTI structural damage: a pilot study.

Participants: Giulia Lioi, Mathis Fleury, Christian Barillot, Isabelle Bonan.

Recent studies have shown the potential of neurofeedback (NF) for motor rehabilitation after stroke. The majority of these NF approaches have relied solely on one imaging technique: mostly on EEG recordings. Recent study have gone further, revealing the potential of integrating complementary techniques such as EEG and fMRI to achieve a more specific regulation. In this exploratory work, multisession bimodal EEG-fMRI NF for upper limb motor recovery was tested in four stroke patients.The feasibility of the NF training was investigated with respect to the integrity of the corticospinal tract (CST), a well-established predictor of the potential for clinical improvement. Results indicated that patients exhibiting a high degree of integrity of the ipsilesional CST showed significant increased activation of the ipsilesional M1 at the end of the training ($p<<0.001$, Wilcoxon test). These preliminary findings confirm the critical role of the CST integrity for stroke motor recovery and indicate that this is importantly related also to functional brain regulation of the ipsilesional motor cortex [43].

7.2.2. Multiple sclerosis

7.2.2.1. Tissue microstructure information from T2 relaxometry and diffusion MRI can identify multiple sclerosis (MS) lesions undergoing blood-brain barrier breakdown (BBB)

Participants: Olivier Commowick, Christian Barillot.

Gadolinium-based contrast agents (GBCA) play a critical role in identifying MS lesions undergoing BBB which is of high clinical importance. However, repeated use of GBCAs over a long period of time and the risks associated with administering it to patients with renal complications has mandated for greater caution in its usage. In this work we explored the plausibility of identifying MS lesions undergoing BBB from tissue microstructure information obtained from T2 relaxometry and dMRI data. We also proposed a framework to predict MS lesions undergoing BBB using the tissue microstructure information and demonstrated its potential on a test case [26].

7.2.2.2. Neural basis of irony in patients with Multiple Sclerosis: an exploratory fMRI study
Participants: Quentin Duché, Élise Bannier.

Irony is a form of non-literal language that is characterized by the opposition between the literal meaning of a statement and the message that the speaker wishes to convey. Knowledge about the neural bases of non-literal language has largely developed in recent years from injury studies or more recently through data from functional imaging studies. Multiple sclerosis (MS) is a neurodegenerative disease that, in addition to cognitive dysfunction, results in variable impairment of theory of mind and non-literal language skills. This work aims at exploring neural basis underpinning the comprehension of irony in MS patients compared to a group of healthy subjects. The results suggest that multiple sclerosis patients require higher left hemisphere resources than healthy controls to understand irony [32].

This work is done in collaboration with by Florian Chapelain (Pôle Saint Hélier), Philippe Gallien (Pôle Saint Hélier) and Virginie Dardier (Université Rennes 2).

7.2.2.3. Joint assessment of brain and spinal cord motor tract damage in patients with early relapsing remitting multiple sclerosis (RRMS): predominant impact of spinal cord lesions on motor function
Participants: Benoit Combès, Élise Bannier, Haykel Snoussi, Jean-Christophe Ferré, Christian Barillot.

The effect of structural multiple sclerosis damage to the corticospinal tract (CST) has been separately evaluated in the brain and spinal cord (SC), even though a cumulative impact is suspected. In this work, we evaluated CST damages on both the cortex and cervical SC, and examine their relative associations with motor function, measured both clinically and by electrophysiology. This study highlights the major contribution of SC lesions to CST damage and motor function abnormalities [8].

This work was done in collaboration with Anne Kerbrat (Neuropoly Montréal) and Raphael Chouteau (CHU Rennes).

7.2.2.4. Spatial distribution of multiple sclerosis lesions in the cervical spinal cord
Participants: Élise Bannier, Gilles Edan.

Spinal cord lesions detected on MRI hold important diagnostic and prognostic value for multiple sclerosis. Our aim was to explore the spatial distribution of multiple sclerosis lesions in the cervical spinal cord, with respect to clinical status. We included 642 suspected or confirmed multiple sclerosis patients (31 clinically isolated syndrome, and 416 relapsing-remitting, 84 secondary progressive, and 73 primary progressive multiple sclerosis) from 13 clinical sites. With an automatic publicly-available analysis pipeline we produced voxelwise lesion frequency maps to identify predilection sites in various patient groups characterized by clinical subtype, Expanded Disability Status Scale score and disease duration. We also measured absolute and normalized lesion volumes in several regions of interest using an atlas-based approach, and evaluated differences within and between groups. The lateral funiculi were more frequently affected by lesions in progressive subtypes than in relapsing in voxelwise analysis ($P < 0.001$), which was further confirmed by absolute and normalized lesion volumes ($P < 0.01$). The central cord area was more often affected by lesions in primary progressive than relapse-remitting patients ($P < 0.001$). Between white and grey matter, the absolute lesion volume in the white matter was greater than in the grey matter in all phenotypes ($P < 0.001$); however when normalizing by each region, normalized lesion volumes were comparable between white and grey matter in primary progressive patients. Lesions appearing in the lateral funiculi and central cord area were significantly correlated with Expanded Disability Status Scale score ($P < 0.001$). High lesion frequencies were observed in patients with a more aggressive disease course, rather than long disease duration. Lesions located in the lateral funiculi and central cord area of the cervical spine may influence clinical status in multiple sclerosis. This work shows the added value of cervical spine lesions, and provides an avenue for evaluating the distribution of spinal cord lesions in various patient groups [14].

This work was done in collaboration with Julien Cohen-Adad (Neuropoly, Montreal) and Anne Kerbrat (Neuropoly Montréal).

7.2.2.5. Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks

Participants: Élise Bannier, Gilles Edan.

The goal of this study was to develop a fully-automatic framework - robust to variability in both image parameters and clinical condition - for segmentation of the spinal cord and intramedullary MS lesions from conventional MRI data of MS and non-MS cases. Scans of 1042 subjects (459 healthy controls, 471 MS patients, and 112 with other spinal pathologies) were included in this multi-site study (n=30). Data spanned three contrasts (T1-, T2-, and $T2_*$-weighted) for a total of 1943vol and featured large heterogeneity in terms of resolution, orientation, coverage, and clinical conditions. The proposed cord and lesion automatic segmentation approach is based on a sequence of two Convolutional Neural Networks (CNNs). CNNs were trained independently with the Dice loss. When compared against manual segmentation, our CNN-based approach showed a median Dice of 95% vs. 88% for PropSeg ($p \leq 0.05$), a state-of-the-art spinal cord segmentation method. Regarding lesion segmentation on MS data, our framework provided a Dice of 60%, a relative volume difference of -15%, and a lesion-wise detection sensitivity and precision of 83% and 77%, respectively. In this study, we introduce a robust method to segment the spinal cord and intramedullary MS lesions on a variety of MRI contrasts. The proposed framework is open-source and readily available in the Spinal Cord Toolbox.

This work was done in collaboration with Julien Cohen-Adad (Neuropoly, Montreal) and Anne Kerbrat (Neuropoly Montréal).

7.2.3. Arterial Spin Labeling in pediatric populations

7.2.3.1. Changes in brain perfusion in successive arterial spin labeling MRI scans in neonates with hypoxic-ischemic encephalopathy

Participants: Maia Proisy, Isabelle Corouge, Antoine Legouhy, Christian Barillot, Jean-Christophe Ferré.

The primary objective of this study was to evaluate changes in cerebral blood flow (CBF) using arterial spin labeling MRI between day 4 of life (DOL4) and day 11 of life (DOL11) in neonates with hypoxic-ischemic encephalopathy (HIE) treated with hypothermia. The secondary objectives were to compare CBF values between the different regions of interest (ROIs) and between infants with ischemic lesions on MRI and infants with normal MRI findings. We prospectively included all consecutive neonates with HIE admitted to the neonatal intensive care unit of our institution who were eligible for therapeutic hypothermia. Each neonate systematically underwent two MRI examinations as close as possible to day 4 (early MRI) and day 11 (late MRI) of life. We proposed an innovative processing pipeline for morphological and ASL data suited to neonates that enable automated segmentation to obtain CBF values over ROIs. We evaluated CBF on two successive scanswithin thefirst 15 days of life in the same subjects. ASL imaging in asphyxiated neonates seems more relevant when used relatively early, in the first days of life. The correlation of intra-subject changes in cerebral perfusion between early and late MRI with neurodevelopmental outcome warrants investigation in a larger cohort, to determine whether the CBF pattern change can provide prognostic information beyond that provided by visible structural abnormalities on conventional MRI [18], [47].

7.2.4. Cerebral blood flow in sickle cell populations

7.2.4.1. White matter has impaired resting oxygen delivery in sickle cell patients

Participant: Julie Coloigner.

Although modern medical management has lowered overt stroke occurrence in patients with sickle cell disease (SCD), progressive white matter (WM) damage remains common. It is known that cerebral blood flow (CBF) increases to compensate for anemia, but sufficiency of cerebral oxygen delivery, especially in the WM, has not been systematically investigated. Cerebral perfusion was measured by arterial spin labeling in 32 SCD patients (age range: 10-42 years old, 14 males, 7 with hemoglobin SC, 25 hemoglobin SS) and 25 age and race-matched healthy controls (age range: 15-45 years old, 10 males, 12 with hemoglobin AS, 13 hemoglobin AA); 8/24 SCD patients were receiving regular blood transfusions and 14/24 non-transfused SCD patients were taking hydroxyurea. Imaging data from control subjects were used to calculate maps for CBF and oxygen delivery in SCD patients and their T-score maps. Whole brain CBF was increased in SCD patients with a mean T-score of 0.5 and correlated with lactate dehydrogenase ($r2 = 0.58$, $P < 0.0001$). When corrected for oxygen content and arterial saturation, whole brain and gray matter (GM) oxygen delivery were normal in SCD, but WM oxygen delivery was 35% lower than in controls. Age and hematocrit were the strongest predictors for WM CBF and oxygen delivery in patients with SCD. There was spatial co-localization between regions of low oxygen delivery and WM hyperintensities on T2 FLAIR imaging. To conclude, oxygen delivery is preserved in the GM of SCD patients, but is decreased throughout the WM, particularly in areas prone to WM silent strokes [7].

This work was done in collaboration with Natasha Leporé and her team, Children's hospital Los Angeles, University of Southern California, USA.

7.2.5. Alzheimer disease

7.2.5.1. Abnormal fMRI response in sub-hippocampal structures: how prior knowledge impairs memory in AD

Participants: Quentin Duché, Pierre-Yves Jonin.

Early Alzheimer's disease typically impairs associative learning abilities, up to 18 years before dementia. Importantly, patients' concerns refer to their daily routine, meaning that they lack associative memory for highly familiar stimuli. However, most of the tests involve much less familiar stimuli (e.g. isolated words). It follows that we ignore whether prior knowledge about memoranda alters memory formation and its neural correlates in Alzheimer's Disease. Here, we aimed at manipulating prior knowledge available at encoding and repetition to investigate whether prior knowledge could alter the neural underpinnings of associative encoding, in a way sensitive to early AD. The results suggest that distinct forms of prior knowledge may drive partly non-overlapping brain networks at encoding, and in turn these regions differentially contribute to successful memory formation. Thus, our finding that sub-hippocampal, not hippocampal, activation underlie the inability of the patients to benefit remote prior knowledge in new learning opens perspectives for further diagnostic and prognostic markers development [37].

7.2.5.2. Learning what you know: how prior knowledge impairs new associative learning in early AD.

Participants: Pierre-Yves Jonin, Quentin Duché, Élise Bannier, Isabelle Corouge, Jean-Christophe Ferré, Christian Barillot.

While associative memory impairment is a core feature of prodromal Alzheimer's Disease (AD), whether prior knowledge affects associative learning is largely overlooked. Stimuli repetition yields suppression or enhancement of the BOLD signal, allowing the functional mapping of brain networks. We addressed the role of prior knowledge in associative encoding by manipulating repetition and familiarity of the memoranda in a subsequent memory fMRI study design. 17 patients with prodromal AD (AD-MCI) and 19 Controls learned face-scene associations presented twice in the scanner. Pre-experimental knowledge trials (PEK) involved famous faces while in Experimental Knowledge trials (EK), unknown faces familiarized before scanning were used. Study events were sorted as associative hits, associative misses or misses after a recognition test outside the scanner. We computed the Repetition X Prior knowledge interaction contrast to test whether the encoding networks differed along with prior knowledge, then looked for subsequent associative memory effects in the resulting clusters. PEK and EK yielded similar associative memory performance in AD-MCI, while PEK increased associative memory by 28% in Controls. Repetition effects were modulated by Prior knowledge in Controls, but AD-MCI showed aberrant repetition effects. Subsequent memory effects were observed only in Controls for PEK in the right subhippocampal structures. By contrast, in both groups, EK triggered a subsequent memory effect in the right hippocampus. Provided that tau pathology starts within anterior subhippocampal regions in early AD, our findings that subhippocampal, not hippocampal, involvement underlies the inability of the patients to benefit from PEK open innovative clinical and research perspectives [38].

7.2.6. Depression

7.2.6.1. White matter abnormalities in depression: a categorical and phenotypic diffusion MRI study.

Participants: Julie Coloigner, Olivier Commowick, Isabelle Corouge, Christian Barillot.

Mood depressive disorder is one of the most disabling chronic diseases with a high rate of everyday life disability that affects 350 million people around the world. Recent advances in neuroimaging have reported widespread structural abnormalities, suggesting a dysfunctional frontal-limbic circuit involved in the pathophysiological mechanisms of depression. However, a variety of different white matter regions has been highlighted and these results lack reproducibility of such categorical-based biomarkers. These inconsistent results might be attributed to various factors: actual categorical definition of depression as well as clinical phenotype variability. In this study, we 1/ examined WM changes in a large cohort (114 patients) compared to a healthy control group and 2/ sought to identify specific WM alterations in relation to specific depressive phenotypes such as anhedonia (i.e. lack of pleasure), anxiety and psychomotor retardation –three core symptoms involved in depression. Consistent with previous studies, reduced white matter was observed in the genu of the corpus callosum extending to the inferior fasciculus and posterior thalamic radiation, confirming a frontal-limbic circuit abnormality. Our analysis also reported other patterns of increased fractional anisotropy and axial diffusivity as well as decreased apparent diffusion coefficient and radial diffusivity in the splenium of the corpus callosum and posterior limb of the internal capsule. Moreover, a positive correlation between FA and anhedonia was found in the superior longitudinal fasciculus as well as a negative correlation in the cingulum. Then, the analysis of the anxiety and diffusion metric revealed that increased anxiety was associated with greater FA values in genu and splenium of corpus callosum, anterior corona radiata and posterior thalamic radiation. Finally, the motor retardation analysis showed a correlation between increased Widlöcher depressive retardation scale scores and reduced FA in the body and genu of the corpus callosum, fornix, and superior striatum. Through this twofold approach (categorical and phenotypic), this study has underlined the need to move forward to a symptom-based research area of biomarkers, which help to understand the pathophysiology of mood depressive disorders and to stratify precise phenotypes of depression with targeted therapeutic strategies [9]. This work was done with Centre Hospitalier Guillaume Régnier, Academic Psychiatry Department, 35703 Rennes, France.

7.2.6.2. Structural connectivity analysis in treatment-resistant depression

Participants: Julie Coloigner, Isabelle Corouge, Christian Barillot.

Depressive disorder is characterized by a profound dysregulation of affect and mood as well as additional abnormalities including cognitive dysfunction, insomnia, fatigue and emotional disturbance. Converging evidence shows that a dysfunction in prefrontal-subcortical circuits is associated with depressive state. However, the process of treatment resistance was poorly studied. One study of functional magnetic resonance imaging has reported more disrupted connectivity in prefrontal areas and in thalamus for resistant (R) group (Lui et al., 2011). These observations suggest a modification of functional connectivity in the prefrontal-subcortical circuits in the R patients. Using graph theory-based analysis, we examined white matter changes in the organization of networks in R patients compared with non-resistant (NR) group. We revealed 15 areas with significant density differences in R patients compared to NR subjects. The NR depression seems associated with decreased connectivity among distributed limbic areas, particularly in the ACC and in basal ganglia. However, the R patients exhibit a reduced connectivity in anterior limb of internal capsule and genu of corpus callosum compared with NR patients. Combined with previous studies, which described a widespread disruption in prefrontal-subcortical networks, this result suggests a more important connectivity decrease in the frontal cortex, as well as a smaller reduction in the limbic circuit for the patients with pejorative outcome. These results were consistent with connectivity studies, which suggested that the degree of disruption could influence the resistance severity and that two distinct networks could be implicated in NR end R depression. [27].

7.2.7. Prenatal exposure

7.2.7.1. Prenatal exposure to glycol ethers and motor inhibition function evaluated by functional MRI at the age of 10 to 12years in the PELAGIE mother-child cohort

Participants: Élise Bannier, Christian Barillot.

Pregnant women are ubiquitously exposed to organic solvents, such as glycol ethers. Several studies suggest potential developmental neurotoxicity following exposure to glycol ethers with a lack of clarity of possible brain mechanisms. We investigated the association between urinary levels of glycol ethers of women during early pregnancy and motor inhibition function of their 10- to 12-year-old children by behavioral assessment and brain MR imaging. Prenatal urinary levels of two glycol ether metabolites were associated with poorer Go/No-Go task performance. Differential activations were observed in the brain motor inhibition network in relation with successful inhibition, but not with cognitive demand. Nevertheless, there is no consistence between performance indicators and cerebral activity results. Other studies are highly necessary given the ubiquity of glycol ether exposure [5].

This work is done in collaboration with Fabienne Pelé and Cécile Chevrier (IRSET). Anne Claire Binter defended her PhD in December 2019 supervised by Fabienne Pelé, Cécile Chevrier and Élise Bannier. t

7.2.7.2. Effect of prenatal organic solvent exposure on structural connectivity at childhood

Participants: Julie Coloigner, Élise Bannier, Jean-Christophe Ferré, Christian Barillot.

Glycol ethers are part of organic solvents. They are used in industry and at home during manufacturing or usage of products such as paints, cleaning agents and cosmetics. The specific detection of subtle, low-dose effects of early-life exposure to these solvents on neuropsychological performance in children is a trendy subject of investigation. Neuroimaging allows looking into brain function and identifying different cerebral connections that may be affected by these neurotoxicants. In this paper, we investigated the specific effects of prenatal low-level exposure to different glycol ethers, on brain development of children between 10 and 12 years old. Based on previous studies suggesting cognitive disabilities in the attention, inhibition and working memory, we proposed a structural connectivity analysis using graph theory restricted to the regions involved in these functions. Our results suggest a possible relationship between the attention, working memory and inhibition and prenatal exposure to specific glycol ethers, such as ethoxyacetic acid, ethoxyethoxyacetic acid and 2-butoxyacetic acid [28].

7.2.8. Cognitive food-choice task

7.2.8.1. Implemenation of a new food picture database in the context of fMRI and visual cognitive food-choice task in healthy volunteers

Participant: Élise Bannier.

This pilot study aimed at implementing a new food picture database in the context of functional magnetic resonance imaging (fMRI) cognitive food-choice task, with an internal conflict or not, in healthy normal-weight adults. The fMRI analyses showed that the different liking foods (i.e. foods with different hedonic appraisals) condition elicited the activation of dorsal anterior cingulate cortex, involved in internal conflict monitoring, whereas similar liking (ie, foods with similar hedonic appraisals) condition did not, and that low-energy (LE) food choice involved high-level cognitive processes with higher activation of the hippocampus (HPC) and fusiform gyrus compared to high-energy (HE) food choice. Overall, this pilot study validated the use of the food picture database and fMRI-based procedure assessing decision-making processing during a food choice cognitive task with and without internal conflict[15].

This work was done in collaboration with Yentl Gautier, Paul Meurice, Yann Serrand, Nicolas Coquery Romain Moirand and David Val-Laillet from the NuMeCan Institute (Nutrition Metabolisms Cancer, UMR 1241, Inserm - Université de Rennes 1) and INRA.

7.3. Research axis 3: Management of Information in Neuroimaging

In the context of population imaging, we have made progress in three main areas this year. First we were involved in the development of infrastructures for open science with OpenAIRE, we also participated in the collaborative definition of standards that will ensure that infrastructures remain interoperable. Finally, we started a research new axis looking at how variations in analytical pipelines impact neuroimaging results (i.e. analytic variability).

7.3.1. Infrastructures

7.3.1.1. Open research: linking the bits and pieces with OpenAIRE-connect

Participants: Camille Maumet, Christian Barillot, Xavier Rolland.

Open research is growing in neuroimaging. The community — supported by funders who want best use of public funding but also by the general public who wants more transparent and participatory research practices — is constantly expanding online resources including: data, code, materials , tutorials, etc. This trend will likely amplify in the future and is also observed in other areas of experimental sciences. Open resources are typically deposited in dedicated repositories that are tailored to a particular type of artefact. While this is best practice, it makes it difficult to get the big picture: artefacts are scattered across the web in a multitude of databases. Although one could claim that the publication is here to link all related artefacts together, it is not machine-readable and does not me toallow searching for artefacts using filters (e.g. all datasets created in relation with a given funder). We presented OpenAIRE-connect, an overlay platform that links together research resources stored on the web: https://beta.ni.openaire.eu/ [45].

This work was done in collaboration with Dr. Sorina Caramasu-Pop and Axel Bonnet from Creatis in Lyon and with collaborators of the OpenAIRE-Connect project.

7.3.2. Standardisation and interoperability

7.3.2.1. The best of both worlds: using semantic web with JSON-LD. An example with NIDM-Results & Datalad

Participant: Camille Maumet.

The Neuroimaging data model (NIDM-Results) provides a harmonised representation for fMRI results reporting using Semantic Web technologies. While those technologies are particularly well suited for aggregation across complex datasets, using them can be costly in terms of initial development time to generate and read the corresponding serialisations. While the technology is machine accessible, it can be difficult to comprehend by humans. This hinders adoption by scientific communities and by software developers used to more-lightweight data-exchange formats, such as JSON. JSON-LD: a JSON representation for semantic graphs (“JSON-LD 1.1” n.d.) was created to address this limitation and recent extensions to the specification allow creating JSON-LD documents that are structured more similar to simple JSON. This representation is simultaneously readable by a large number of JSON-based applications and by Semantic Web tools. Here we review our work on building a JSON-LD representation for NIDM-Results data and exposing it to Datalad, a data-management tool suitable for neuroimaging datasets with built-in support for metadata extraction and search [44].

This work was done in collaboration with Prof. Michael Hanke from Institute of Neuroscience and Medicine in Julich and with members of the INCF.

7.3.2.2. Tools for FAIR Neuroimaging Experiment Metadata Annotation with NIDM Experiment

Participant: Camille Maumet.

Acceleration of scientific discovery relies on our ability to effectively use data acquired by consortiums and/or across multiple domains to generate robust and replicable findings. Efficient use of existing data relies on metadata being FAIR1 - Findable, Accessible, Interoperable and Reusable. Typically, data are shared using formats appropriate for the specific data types with little contextual information. Therefore, scientists looking to reuse data must contend with data originating from multiple sources, lacking complete acquisition information and often basic participant information (e.g. sex, age). What is required is a rich metadata standard that allows annotation of participant and data information throughout the experiment workflow, thereby allowing consumers easy discovery of suitable data. The Neuroimaging Data Model (NIDM)2 is an ongoing effort to represent, in a single core technology, the different components of a research activity, their relations, and derived data provenance3. NIDM-Experiment (NIDM-E) is focused on experiment design, source data descriptions, and information on the participants and acquisition information. In this work we report on annotation tools developed as part of the PyNIDM4 application programming interface (API) and their application to annotating and extending the BIDS5 versions of ADHD2006 and ABIDE7 datasets hosted in DataLad[40].

This work was led by Dr David Keator from UCI Irvine and done in collaboration with members of the INCF.

7.3.3. Quantifying analytic variability

7.3.3.1. Exploring the impact of analysis software on task fMRI results

Participant: Camille Maumet.

A wealth of analysis tools are available to fMRI researchers in order to extract patterns of task variation and, ultimately, understand cognitive function. However, this 'methodological plurality' comes with a drawback. While conceptually similar, two different analysis pipelines applied on the same dataset may not produce the same scientific results. Differences in methods, implementations across software packages, and even operating systems or software versions all contribute to this variability. Consequently, attention in the field has recently been directed to reproducibility and data sharing. Neuroimaging is currently experiencing a surge in initiatives to improve research practices and ensure that all conclusions inferred from an fMRI study are replicable. In this work, our goal is to understand how choice of software package impacts on analysis results. We use publically shared data from three published task fMRI neuroimaging studies, reanalyzing each study using the three main neuroimaging software packages, AFNI, FSL and SPM, using parametric and nonparametric inference. We obtain all information on how to process, analyze, and model each dataset from the publications. We make quantitative and qualitative comparisons between our replications to gauge the scale of variability in our results and assess the fundamental differences between each software package. While qualitatively we find broad similarities between packages, we also discover marked differences, such as Dice similarity coefficients ranging from 0.000-0.743 in comparisons of thresholded statistic maps between software. We discuss the challenges involved in trying to reanalyse the published studies, and highlight our own efforts to make this research reproducible [6].

This work was done in collaboration with Alexander Bowring and Prof. Thomas Nichols from the Oxford Big Data Institute in the UK.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Siemens

Participants: Elise Bannier, Christian Barillot, Emmanuel Caruyer, Olivier Commowick, Isabelle Corouge, Jean-Christophe Ferré, Jean-Yves Gauvrit.

In the context of the Neurinfo imaging platform, a master research agreement between Siemens SAS - Healthcare and University of Rennes 1 defines the terms of the collaboration between Siemens, Empenn and the Neurinfo platform. Relying on 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. The MR Diffusion pulse sequence source code was modified in collaboration with our Siemens clinical scientist as part of our Master Research Agreement, Marc Lapert, in order to play arbitrary gradient waveforms. This was done on the Syngo VB17 software version and again VE11C (nearly finished).

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Défis scientifiques 2019 of University of Rennes 1: Compensating analytic variability for a better use of open data (2019, 6500€).

Participant: Camille Maumet.

In neuroimaging, open data are now well developed with hundred of thousands of images available for the research community. However, those data are still mainly studied in isolation, limiting the potential for new discoveries. Here we focus our efforts on developing neuroinformatics standards and algorithms that will support publication and combination of open datasets.

9.1.2. Region Bretagne: project VARANASI

Participants: Christian Barillot, Camille Maumet, Xavier Rolland.

Thanks to the development of open science practices, more and more public datasets are available to the research community. In the field of brain imaging, these data, combined, bring a critical increase in sample size, necessary to build robust models of the typical and atypical brain. However, in order to build valid inferences on these data, we need to take into account their heterogeneity. Variability can arise due to multiple factors such as: differences in imaging instruments, in acquisitions protocols and even, in post-processing pipelines. In particular, the expansion of open source machine learning workflows creates a multitude of possible outputs out of the same dataset. The variations induced by this methodological plurality can be reFerréd to as 'analytic variability' which will be the focus of the thesis funded in half by region Bretagne. The thesis of Xavier Rolland (2018-2021) will address two challenges: 1) How to combine neuroimaging data generated by different analysis pipelines? 2) How to publish neuroimages with an adequate level of metadata to enable their reuse? Methodological developments will combine machine learning techniques with methods from knowledge representation.

9.2. National Initiatives

9.2.1. Projet Fondation de France: PERINE: 99k€ (33 k€ for IRM acquisition and 22 k€ for image analysis) for 2011-2021

Participants: Élise Bannier, Isabelle Corouge, Julie Coloigner, Maia Proisy, Jean-Christophe Ferré, Christian Barillot.

The PELAGIE cohort evaluates the effect of prenatal exposure to neurotoxicants on child development. Following previous studies, the PERINE study focuses on the assessment of brain developement at 10-12 years old using MRI (ASL, Diffusion imaging, Working memory as well as motor inhibition BOLD fMRI together with neuropsychological tests). A total of 101 children were included. A PhD of Anne-Claire Binter was defended in December 2019 linking epidemiology with functional imaging during a GoNoGo task and neuropsychological scores. This work is done in collaboration with Fabienne Pele´ and Ce´cile Chevrier (IRSET).

9.2.2. Fondation de l'Avenir: EPMR-MA

Participants: Pierre-Yves Jonin, Élise Bannier, Christian Barillot, Quentin Duché.

Recognition memory refers to our ability to discriminate between previously experienced vs. unexperienced stimuli. It is impaired very early in the course of Alzheimer's Disease (AD), both regarding behavioural performance and related brain activity. When the memoranda is associated or with existing knowledge, subsequent memory increases in healthy subjects. Moreover, existing knowledge related to prior exposures may alter the brain network underlying successful memory formation. While much is known regarding the brain substrates of recognition memory in early AD, little is known about the impact of prior exposure. Yet, this factor could both enhance memory formation in patients, and highlight a pattern of memory impairments and related brain activity that might accurately discriminates between early AD, before dementia, and healthy aging. The present task-based fMRI study aims at assessing the influence of prior exposures on recognition memory and its neural underpinnings in patients with Mild Cognitive Impairment due to AD. Inclusions were performed between 2016 and 2017 and data analysis is ongoing.

9.2.3. Projet Fondation de France: Connectivity of the amygdala in depression: (PI: M.-L. Paillère Martinot, Paris Descartes University), €200k for 2018-2021

Participants: Christian Barillot, Olivier Commowick, Emmanuel Caruyer, Julie Coloigner.

The onset of depression in teenagers and young adults increases the risk to develop a drug-resistant depression in the adulthood. This project aims at evaluating the role of early changes in the microstructure and connectivity of the amygdala. Using a cohort of drug-resistant patients (N=30), non drug-resistant patients (N=30) and controls (N=30), the aim is to identify imaging biomarkers of the pathology and to compare these with emotional and cognitive phenotypes in this population, searching for early differences in the development of the amygdala connectivity.

9.2.4. CNRS-Inserm Défi Santé numérique – AAP 2019: Imagerie Multimodale de l'Amygdale limbiquE pour le pronostic de la Dépression (IMpAirED): 19k€ for 2019

Participants: Julie Coloigner, Olivier Commowick, Élise Bannier, Emmanuel Caruyer, Christian Barillot.

This grant is an extension of the Projet Fondation de France: Connectivity of the amygdala in depression.

In order to identify early features of this depression disease, the aim of this project is to develop multimodal modeling of the limb amygdala and its network from MR imaging combining activation and rest functional imaging and MR brain microstructure imaging quantitative (diffusion and relaxometry). The development of this model will allow us to define three imaging biotypes corresponding to depressed adult patients responding to antidepressant treatments, depressed resistant patients and controls. These multimodal imaging biomarkers will be used to stratifie a large longitudinal cohort of young adults into three sub-groupsgroups, in order to retrospectively identify early differences in development trajectories of amygdala.

Inclusions of the patients will begin in early 2020.

9.2.5. ANR "MAIA", generic projects program: €150k for 2016-2019 (PI: F. Rousseau, IMT Atlantique, Brest)

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 is 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 is the design of such software tools for medical image / signal analysis, actually operational in clinical routine, and freely available. Academic researchers and industrial partners are working in close collaboration to reach that ambitious goal.

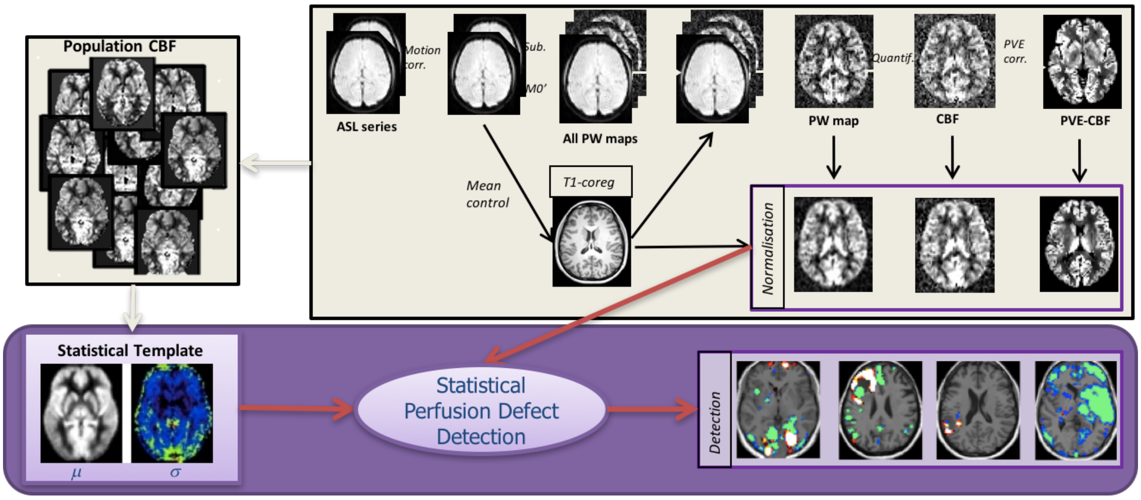


Figure 2. Processing workflow for quantification of Arterial Spin Labelling Cerebral Blood Flow with detection of abnormal perfusion

9.2.6. Fondation pour la recherche médicale (FRM) - Project Hybrid EEG/IRM Neurofeedback for rehabilitation of brain pathologies: 370k€ (2017-2021)

Participants: Élise Bannier, 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: Empenn U1228, HYBRID and PANAMA Teams from Inria/Irisa Rennes and EA 4712 team from University of Rennes I.

9.2.7. PHRC EMISEP: Evaluation of early spinal cord injury and late physical disability in Relapsing Remitting Multiple Sclerosis: €200k for 2016-2019

Participants: Élise Bannier, Christian Barillot, Emmanuel Caruyer, Benoit Combès, Olivier Commowick, Gilles Edan, Jean-Christophe Ferré, Haykel Snoussi.

Multiple Sclerosis (MS) is the most frequent acquired neurological disease affecting young adults (1 over 1000 inhabitants in France) and leading to impairment. Early and well adapted treatment is essential for patients presenting aggressive forms of MS. This PHRC (Programme hospitalier de recherche clinique) project focuses on physical impairment and especially on the ability to walk. Several studies, whether epidemiologic or based on brain MRI, have shown that several factors are 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) Expanded Disability Status Scale 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 two 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. Haykel Snoussi defended his PhD Thesis on diffusion imaging in the spinal cord starting with distortion correction.

B. Combe`s 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.

9.2.8. MS-TRACTS (ARSEP and COREC funding): Estimating the impact of multiple sclerosis lesions in motor and proprioceptive tracts, from the brain to the thoracic spinal cord, on their functions, assessed from clinical tests and electrophysiological measurements: 45k€ (2019-2021).

Participants: Élise Bannier, Benoit Combès.

Previous 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, clinical relapses, number of focal lesions on baseline MRI. However, these factors only partially explain physical disability progression, preventing their use at the individual level. The access to advanced brain and cord MR images, the development of associated processing tools combined. We hypothesize that a fine assessment of damage on specific networks, from the brain to the thoracic cord, offers a relevant biomarker of disability progression in MS. Such damage assessments must take into account both lesion location, assessed on structural brain and cord MR images and lesion severity, assessed using quantitative MR images. We propose to test this hypothesis by combining assessments of lesion location and severity on corticospinal and proprioceptive tracts from the brain to the thoracic cord with clinical and electrophysiological measurements. This study includes two French centers (Rennes, Marseille) and includes a total of 60 patients. The expected outcome is to obtain early biomarkers of physical impairment evolution in RRMS patients, first treated with immunomodulatory treatment. The long-term goal is to provide the clinician with biomarkers able to anticipate therapeutic decisions and support the switch to alternative more aggressive treatment.

9.2.9. PIA projects

9.2.9.1. The HEMISFER Project: (€400k for 2017-2019)

Participants: Élise Bannier, Isabelle Bonan, Isabelle Corouge, Claire Cury, Jean-Christophe Ferré, Jean-Yves Gauvrit, Pierre Maurel, Christian Barillot.

The HEMISFER project ("Hybrid Eeg-MrI and Simultaneous neuro-FEedback for brain Rehabilitation") is conducted at Inria Rennes with the support of the Labex "CominLabs" [0]. The goal of HEMISFER is to make

[0] https://iww.inria.fr/cominlabs-newsletter/april-2013-four-projects-selected/#hemisfer

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 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, etc.). This project involves 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 benefits from the research 3T MRI and MRI-compatible EEG systems provided by the NeurInfo in-vivo neuroimaging platform on which these new research protocols aree set up. A budget of 500K€ is provided by CominLabs to support this project (through experimental designs, PhDs, post-docs and expert engineers).

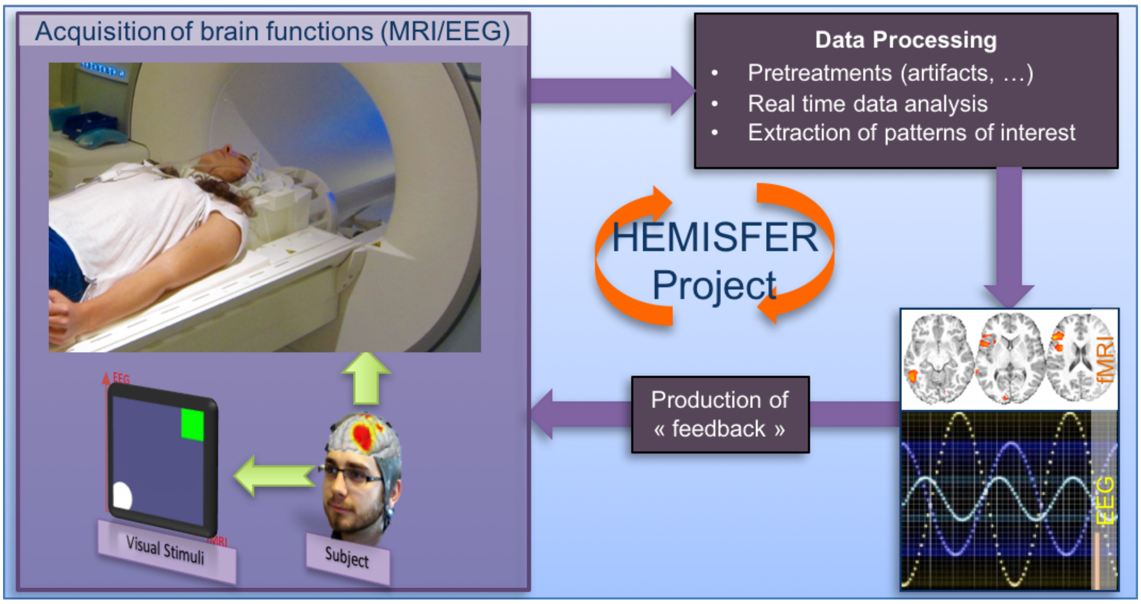


Figure 3. Principle of the Hemisfer project.

9.2.9.2. France Life Imaging (FLI): 2012-2023, €2000k (phase 1) + €1200k (phase 2)

Participants: Christian Barillot, Olivier Commowick.

France Life Imaging (FLI) is a large-scale research infrastructure project to establish a coordinated and harmonized network of biomedical imaging in France. This project was 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 built by a consortium of teams that contribute to the construction of a network for data storage and information processing. Instead of building yet other dedicated facilities, the IAM node use already existing data storage and information processing facilities (LaTIM Brest; CREATIS Lyon; CIC-IT Nancy; Empenn U1228 Inria Rennes; CATI CEA Saclay; ICube Strasbourg) that increase their capacities for the FLI infrastructure. Inter-connections and access to services are achieved through a dedicated software platform that is developed based on the expertise gained through successful existing developments. The IAM node has several goals. It is building a versatile facility for data management that inter-connects the data production sites and data processing for which state-of-the-art solutions, hardware and software, are available to infrastructure users. Modular solutions are preferred to accommodate the large variety of modalities acquisitions, scientific problems, data size, and to be adapted for future challenges. Second, it offers the latest development that are made available to image processing research teams. The team Empenn 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 is the technical manager. Apart from the team members, software solutions like MedInria and Shanoir are part of the software platform.

9.2.9.3. OFSEP: €175k for 2017-2019

Participants: Élise Bannier, Christian Barillot, Olivier Commowick, Gilles Edan, Jean-Christophe Ferré, Francesca Galassi.

The French Observatory of Multiple Sclerosis (OFSEP) is one of ten 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 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. FP7 & H2020 Projects

9.3.1.1. OpenAire-Connect

Participants: Christian Barillot, Camille Maumet, Xavier Rolland.

Project title: **OpenAire-Connect**

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 (Empenn, 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 € (120k€ for CNRS)

The OpenAire-Connect H2020 project introduces and implements the concept of Open Science as a Service (OSaaS) on top of the existing OpenAIRE infrastructure, delivering out-of-the-box, on-demand deployable tools. OpenAIRE-Connect adopts an end-user driven approach (via the involvement of five prominent research communities), and enriches the portfolio of OpenAIRE infrastructure production services with a Research Community Dashboard Service and a Catch-All Notification Broker Service. The first offers 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, provides communities with the content and tools they need to effectively evaluate and reproduce science. OpenAIRE-Connect combines dissemination and training with OpenAIRE's powerful NOAD network engaging research communities and content providers in adopting such services. These combined actions 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 Empenn is acting, through CNRS, as the French coordinator to develop the link with the Neuroimaging research community. This is performed in the context of the FLI-IAM national infrastructure.

9.3.1.2. EIT-Health

Participant: Christian Barillot.

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. Empenn 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". Empenn is also concerned by the WG "big data".

9.4. International Initiatives

9.4.1. Inria International Labs

9.4.1.1. MMINCARAV

EPFL-Inria

Associate Team involved in the International Lab:

Title: Multimodal Microstructure-Informed Neuronal Connectivity: Acquisition, Reconstruction, Analysis and Validation

International Partner (Institution - Laboratory - Researcher):

Ecole Polytechnique Fédérale de Lausanne (Switzerland) - Laboratoire de Traitement du Signal 5 - Jean-Philippe Thiran

Start year: 2019

See also: https://team.inria.fr/empenn/research/mmincarav-inria-epfl/

Participants: Emmanuel Caruyer, Olivier Commowick, Julie Coloigner, Élise Bannier and Christian Barillot.

The objectives of this associate team will be to address new scientific challenges related to the use of multimodal magnetic resonance imaging (MRI) to derive microstructure indices and apply them to the measure of brain connectivity. We will focus on 4 aspects of this: first we will develop novel sampling techniques, with the objective to reduce acquisition time for the accurate reconstruction of microstructure indices using diffusion MRI; next we will propose joint T2 relaxometry and diffusion models for the description of microstructure, to take advantage of the complementarity of both modalities in the estimation of microstructure indices; in continuation, we will propose new statistical and network analysis methods using the microstructure-informed connectome, and evaluate its potential to reduce bias and false positives; last we will develop a realistic simulation tool combining a fine macroscopic description of fiber bundles, with a fast and realistic simulator at the mesoscopic scale developed by LTS5.

9.4.1.2. Other projects

Participants: Pierre Maurel, Christian Barillot, Claire Cury.

Gundishapur Program (Partenariat Hubert Curien franco-iranien)

This project is a collaboration between the Empenn team and the Institute of medical science and technologies (Shahid Beheshti university, Iran).

Combining EEG (Electroencephalogram) and fMRI (functional Magnetic Resonance Imaging) shows great promise in helping scientists to better understand the complex function of the brain. It can also be used in understanding the brain dysfunctions or specific behaviors. The integration of these two modalities can provide a good spatio-temporal resolution of the neuronal activities, and therefore, it can bring a good insight on the brain function. EEG is the recording of the electrical activity of the brain through scalp surface electrodes. We are already working in this area through the HEMISFER project, whose goal is to make full use of neurofeedback paradigm by using a coupling model associating functional and metabolic information from Magnetic Resonance Imaging (fMRI) and Electro-encephalography (EEG) to “enhance” the neurofeedback protocol. A former member of our team, Dr. Noorzadeh, has already worked on a part of this project, and is now in IMSAT (Iran). He is our main contact for this collaboration.

This project works on the integration methods, in order to first acquire the simultaneous data of high quality with the minimum possible artifacts, and also on biomedical applications in this regard. One of these applications is the source localization using the multi-modal data. Identifying neuronal sources in both high spatial and temporal resolution can open up a bright way to understand lots of diseases, among which epilepsy is the main one. The epileptic seizures or the inter-ictal discharges are nowadays only detected by EEG, but the origin of the activity is only inferred in terms of brain lobes. This spatial precision can be augmented and the method can be used in the precise detection of the focal points of epilepsy for the pre-surgical evaluations.

9.4.1.3. Informal International Collaborations

- Emmanuel Caruyer collaborates with Alice Bates, research fellow at Australian National University, Canberra, on "Dimensionality sampling for B-tensor encoding in diffusion MRI".
- Camille Maumet collaborates with Prof. Thomas Nichols and his group, NISOx at the Oxford Big Data Institute, with Prof. Jean-Baptiste Poline and his group at McGill University, with Prof. Satrajit Ghosh and his group at MIT, with Dr David Keator at UCI Irvine, with Dr. Karl Helmer at MGH, with Dr Tristan Glatard and his group at Concordia University and with international members of the INCF on neuroimaging data sharing.
- Julie Coloigner collaborates with Prof. Natasha Leporé and Dr. John Wood, Children’s hospital Los Angeles, University Southern California.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- David Kennedy, Professor at University of Massachusetts Medical School, US visited the team on Feb 27 and gave a talk on Repronim: a center for reproducible neuroimaging.
- Natasha Leporé, Professor at Children’s hospital Los Angeles, University of Southern California, US visited the team on April 4-5. She gave a talk on "Understanding pediatric brain anatomy through MRI".
- Jan Petr, Researcher at the HZDR in Dresden visited the team in March 1st, 2019 and give a talk on "Processing ASL data with ExploreASL - technical improvement and clinical applications".

9.5.2. Visits to International Teams

9.5.2.1. Research Stays Abroad

- Corentin Vallée visited Brainnetome center, Institute of Automation, Chinese Academy of Science, Beijing from June 1, 2019 to July 31, 2019; he was awarded a grant for international mobility from the MathSTIC doctoral school.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

10.1.1.1. General Chair, Scientific Chair

- Camille Maumet was chair of the Open Science Special Interst Group of the Organisation of Human Brain Mapping. https://www.humanbrainmapping.org/m/pages.cfm?pageid=3712
- Giulia Lioi was chair of a Scientific Symposium during the conference of Organisation of Human Brain Mapping, 2019 : "Multimodal Neurofeedback: The next generation of neurofeedback for advanced brain self-regulation"
- Brainhack Global Rennes 2019: https://brainhack.irisa.fr, November 14-15, Rennes, France (Julie Coloigner, Emmanuel Caruyer and Raphael Truffet).
- Christian Barillot is member of the Board of Directors of IPMI conference series (Information Processing in Medical Imaging)
- Camille Maumet is a member of the "Comité national pour la Science Ouverte (CoSO)" Working group of open software led by Roberto Di Cosmo.
- Camille Maumet is a member of Aperture, a new publishing plateform, working group “workflows” led by Peter Bandettini.

10.1.1.2. Member of the Organizing Committees

- Brainhack Global Rennes 2018: https://brainhack.irisa.fr, May 25-26, Rennes, France (Julie Coloigner, Emmanuel Caruyer and Raphael Truffet).
- ARSEP workshop on multiple sclerosis, Feb 2, ICM (Élise Bannier).
- Élise Bannier is founding member of the REMI network for mutual aid in MRI studies (https://remi.network/) to facilitate setup and quality control of multicenter studies and share expertise between centers, with two meetings organized this year in Strasbourg and Paris.

10.1.2. Scientific Events: Selection

10.1.2.1. Member of the Conference Program Committees

- Emmanuel Caruyer has been a program committee member of the MICCAI CDMRI workshop since 2013.
- Camille Maumet is a member of programm committee of INCF Neuroinformatics 2019 https://www.neuroinformatics2019.org/, September 1-2, Warsaw, Poland.
- Camille Maumet is member of Program Commitee: GEANT workshop: https://neuroinfo.fr/#!/workshops/geant2019, May 21, Marseille, France

10.1.2.2. Reviewer

- IEEE International Symposium on Biomedical Imaging (Emmanuel Caruyer, Olivier Commowick, Julie Coloigner)
- MICCAI (Olivier Commowick, Francesca Galassi)
- Workshop on Biomedical Image Registration (Olivier Commowick)
- Annual congress of the Organisation of Human Brain Mapping (Camille Maumet)
- ESMRMB (Élise Bannier)

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Camille Maumet is member of Editorial Boards of Neuroinformatics

- Christian Barillot is Chief editor of Frontier in Computer science.

10.1.3.2. Reviewer - Reviewing Activities

- NeuroImage (Emmanuel Caruyer, Camille Maumet, Pierre Maurel)
- Medical Image Analysis (Emmanuel Caruyer)
- IEEE TMI (Olivier Commowick)
- Neuroinformatics (Julie Coloigner)
- Med Image Anal (Olivier Commowick, Emmanuel Caruyer, Christian Barillot)
- Scientific Data (Camille Maumet)

10.1.4. Invited Talks

- Emmanuel Caruyer,"Nouveaux enjeux pour l'acquisition d'IRM de diffusion pour la microstructure" ; Journée Nouvelles Imageries INS2I ; December 2019; Lyon."
- Corentin Vallée, "Arterial Spin Labeling, Resting-State and Acquisition duration", Brainnetome center, Institute of Automation, Chinese Academy of Science, Beijing; June 11, 2019
- Corentin Vallée, "Arterial Spin Labeling, Resting-State and Acquisition duration", Brainnetome center, Institute of biophysics, Chinese Academy of Science, Beijing; June 25, 2019
- Giulia Lioi, "EEG-fMRI integration for Neurofeedback", Diagnosis and International Adaptive Imaging (IADI) Laboratory, Nancy, France; September 2019.
- Giulia Lioi, "Multimodal Neurofeedback for Stroke Rehabilitation", 4eme Journee Nationale Neurofeedback Evaluation and Training (NEXT),Bordeaux, France, May 2019.
- Giulia Lioi "Integrating EEG and fMRI for Neurofeedback", Conference Gen2Bio, Biogenouest, Angers, France, May 2019.
- Pierre-Yves Jonin, "Statistiques appliquées en psychométrie. Journées scientifiques de rééducation cognitive chez l'enfant", Lyon, June 14, 2019.
- Pierre-Yves Jonin, "Les méthodes psychométriques pour le cas unique:", Journée de formation de l'Association des Neuropsychologues en Drôme & Ardèche; Valence, France, November 22, 2019.
- Pierre Maurel, "Medical Image Processing: MRI and Brain", seminar at the Mechatronics Department of ENS-Rennes.
- Élise Bannier, "Multi-center (mono-vendor) longitudinal conventional and quantitative spinal cord MRI at 3 Tesla", Spinal Cord Workshop, UCL, London, January 21, 2019.

10.1.5. Leadership within the Scientific Community

- Emmanuel Caruyer: Member (by selection) of the "Bio Imaging and Signal Processing" (BISP) Technical Committee of the IEEE Signal Processing Society (Emmanuel Caruyer).
- Élise Bannier: Elected Member of the French Society for MRI (SFRMBM), starting in Jan 2020 for 6 years.

10.1.6. Scientific Expertise

- Expertise as a reviewer for "Agence Nationale de la Recherche" (ANR) (Emmanuel Caruyer)
- Élise Bannier: HCERES Evaluation Committees of the CERCO and TONIC units in Toulouse as expert, representing the research support staff.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

L2 informatique: Raphaël Truffet, Génie Logiciel, 24h and Complexity, 18h, L2, ISTIC, France,

L2 informatique: Raphaël Truffet, Génie Logiciel, 40h, L2, ISTIC, France

ENS Rennes: Raphaël Truffet, Algorithmic (20h)

L2 biologie: Antoine Legouhy, Bio-statistques 34h, L2, Univ. Rennes 1, France.

M1 mathématiques: Antoine Legouhy, Optimisation 12h, M1, Univ. Rennes 1, France.

M1 biologie: Antoine Legouhy, statistiscs 12h, M1, Univ. Rennes 1, France.

Master SIBM, M2, University of Angers-Brest-Rennes, France.

L1 INF1: Francesca Galassi, Functional and Immutable Programming (Scala) (TP : 40h), L1, ISTIC University of Rennes 1, France.

L3 bioinformatique: Corentin Vallée, Statistics, 32h, University of Rennes 1, France.

Master SIBM: Benoit Combès, Statistics (TD, TP: 12h), University of Angers-Brest-Rennes:

- Emmanuel Caruyer, “Introduction to diffusion MRI” (Plenary: 3h).
- Benoit Combès, “Méthodes de recalage linéaire et non-linéaires” (Plenary: 6h)
- Quentin Duché, “Traitements des données d’IRM fonctionnelle” (Plenary: 1h).
- Isabelle Corouge, “Bio-marqueurs d’imagerie et IRM metabolique et fonctionnelle” (Plenary: 3h).
- Camille Maumet, “Imaging processing pipelines” (Plenary: 3h).
- Élise Bannier, “IRM fonctionnelle BOLD” (Plenary: 1h).
- Benoit Combès, “Méthodes statistiques pour le traitement d’image” (Plenary: 3h).
- Claire Cury, "Evaluation des performances en imagerie me´dicale" (Plenary: 3h). item Julie Coloigner, "Méthodes d’analyse de la connectivité cérébrale" (Plenary: 3h).

Master 2 "Langage Cognition et Apprentissage: Pierre-Yves Jonin,"Méthodologie clinique & psychométrique de l’étude de cas" (4h); Université de Poitiers, February 2019

DIU Sémiologie des démences: Pierre-Yves Jonin, "Diagnostic précoce de la maladie d’Alzheimer: contribution neuropsychologique" (4h); Université de Caen Basse Normandie, March 2019

Master 2 Neuropsychologie: Pierre-Yves Jonin, "Apports et limites du bilan neuropsychologique à visée diagnostic dans les syndromes démentiels" (3,5h); Université de Savoie à Chambéry, March 2019

Licence 3 Psychologie, UES Handicap: Pierre-Yves Jonin "Déficit mnésique et handicap invisible. Détection en neuropsychologie" (8 h); Université de Rennes 2, February 2019

Master 2 Psychologie spécialité handicap: Pierre-Yves Jonin "Neuropsychologie clinique du sujet âgé" (10 h); Université de Rennes 2, January-April 2019

Medicine (4th & 5th year) UE Psychologie et Neurobiologie: Pierre-Yves Jonin,L’exploration neuropsychologique des maladies neurologiques et psychiatriques. 4ème & 5ème année de médecine (4h), Université de Brest, France.

Licence 3 de Psychologie: Pierre-Yves Jonin, Les syndromes neuropsychologiques. L3, (16h); Université de Rennes 2, France

Master 2 "Troubles de la Cognition et du Langage": Pierre-Yves Jonin, Evaluation neuropsychologique à visée diagnostique. Le cas des syndromes démentiels. Master 2 (2 x 6h) Université de Poitiers, France.

Master 2 Neurosciences Cliniques: Pierre-Yves Jonin, "La mémoire humaine" (3h); Université de Rennes 1, novembre 2019

IUT St-Malo: Mathis Fleury, statistics (64); 2019

Licences 2 & 3 Biologie: Xavier Rolland, "Introduction aux Biostatistiques" 1 (TP sur R: 64h), Université de Rennes 1

ESIR: Pierre Maurel, General image processing (60h), Algorithmics and complexity (60h), Medical imaging (60h), École Supérieure d’Ingénieur de Rennes (ESIR).

ENS Rennes: Pierre Maurel, Introduction to image processing (24h)

10.2.2. Supervision

10.2.2.1. PhD & HdR

HDR: Olivier Commowick, "Compartments imaging for the characterization of brain diseases from quantitative MRI", Univ. Rennes 1, defended on June 19, 2019, jury: Isabelle Bloch - Professeur, Telecom Paristech, Sébastien Ourselin - Professeur, King's College London, Jean-Philippe Thiran - Professeur, Ecole Polytechnique Fédérale de Lausanne, Christian Barillot - DR CNRS, Equipe Empenn, Alexander Leemans - Associate Professor, UMC Utrecht and Xavier Pennec - DR Inria, Equipe Epione[1].

PhD: Anne-Claire Binter, "Effects of prenatal exposure to neurotoxicants on the child's brain function evaluated by cerebral imaging", defended on December 4, 2019, Univ. Rennes, Élise Bannier.

PhD: Cédric Meurée, "Arterial spin labelling : quality control and super-resolution", defended on March 25, 2019, Univ. Rennes, Pierre Maurel, Christian Barillot.

PhD: Haykel Snoussi, "Population imaging and diffusion MRI for characterizing multiple sclerosis in the human spinal cord.", defended on May 2, 2019, Univ. Rennes, Christian Barillot and Emmanuel Caruyer[3].

PhD in progress: Antoine Legouhy, "Longitudinal brain altas creation, application to development studies", CNRS, from Nov 2016, Defence planned for February 2020, 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: Xavier Rolland, "Modeling analytic variability in brain imaging", CNRS, from Oct 2018, Christian Barillot, Camille Maumet.

PhD in progress: Raphaël Truffet, "Compressed sensing for microstructure-enabled diffusion MRI", Univ. Rennes / ENS Rennes, from Oct 2018, Christian Barillot, Emmanuel Caruyer.

PhD in progress: Giovanna Orrù, "Modeling brain structural and functional connectivity ", Univ. Rennes, from Oct 2019, Julie Coloigner and Christian Barillot.

PhD in progress: Stéphanie Leplaideur, "Equilibre de la rééducation par vibrations cervicales, adaptation prismatique et association aux deux techniques, chez des patients cérébro-lésés droits– Protocole AVC POSTIM " , started in October 2018, Isabelle Bonan, Élise Bannier.

10.2.2.2. Other supervisions

Master student SIF (Master mention Informatique, parcours Science informatique): Antoine Ackaouy, "Deep learning for segmentation of MS lesions from multimodal MRI", Apr-Oct 2018, Francesca Galassi.

3-months visit: Gregory Kiar, PhD student, "Analytical Stability in fMRI" April-June 2019, Camille Maumet in collaboration with Elisa Fromont (LACODAM, Inria Rennes / IRISA).

Co-supervised project : Alexander Bowring, "Exploring the Impact of Analysis Software on Task fMRI Results" 2019, Camille Maumet in collaboration with Thomas Nichols (LACODAM, Inria Rennes / IRISA).

Co-supervised project : Freya Acar, Best practices reporting for fMRI meta-analysis 2019, Camille Maumet in collaboration with Beatrijs Moerkerke (LACODAM, Inria Rennes / IRISA).

M1 Physique Médicale student: Marion Boulanger, "Evaluation of QC procedures at Neurinfo and optimisation".

L3 ESIR Ingénierie Biomédicale student: Chloé Mercier, "Implementation of a Java tool to perform MR Quality Control on a Eurospin test object".

10.2.3. Juries

- Emmanuel Caruyer. PhD committee: Yann Bihan-Poudec, Université Claude Bernard Lyon 1; December 17, 2019.
- Olivier Commowick. PhD committee: Junhao Wen, "Structural and microstructural neuroimaging for diagnosis and tracking of neurodegenerative diseases", Sorbonne Université; July 4, 2019, PhD supervisors: Olivier Colliot et Stanley Durrleman.
- Olivier Commowick. MD thesis committee: Charlotte Laurent "Intérêt pronostique de l'IRM en tenseur de diffusion dans les neuropathies optiques inflammatoires", Université de Rennes 1; Octobre 25, 2019.
- Olivier Commowick. PhD committee: Christian El Hajj "Estimation et classification des temps de relaxation multi-exponentiels en IRM. Application aux tissus ve´ge´taux.", Ecole centrale Nantes; December 16, 2019, PhD supervisors: Guylaine Collewet and Maja Musse.

10.3. Popularization

10.3.1. Brain awareness week: “Semaine du Cerveau”

Empenn team took part in Brain awareness week 2019. Giulia Lioi, Claire Cury and Élise Bannier coordinated this event for Empenn/Neurinfo.

- Camille Maumet and Christian Barillot: Talk "" Le cerveau numérique: explorons-le!"". March 2019, Rennes, France.
- Quiz on brain imaging, March 2019, Rennes, France.
- Lab visits for students (3e and 1ere S) about MRI, functional MRI, EEG and neurofeedback, and image processing.
- Pierre-Yves Jonin: Talk "Entre connaissances et souvenirs : voyage en terre mnésique", March 2019, Rennes, France

10.3.2. Journée Science et Musique (JSM)

This event was initiated in 2010 by IRISA. The idea is to offer a scientific and festive immersion in the research and new technologies of the music of neurosciences and sounds to the general public. Free workshops, conferences, demonstrations and concerts are available to all throughout the day (for more detail see https://jsm.irisa.fr/). Claire Cury organized the event this year with Nancy Bertin, PANAMA team.

10.3.3. Sciences en Cour[t]s

This event is a festival of short movies, which offers doctoral students the opportunity to make short films about their thesis work. Raphael Truffet, Antoine Legouhy and Xavier Rolland won the high school award https://www.youtube.com/watch?v=IKgqv-iCwak.

10.3.4. "J'peux pas j'ai informatique"

"J'peux pas j'ai informatique" is an outreach program to welcome secondary school students for a day at Inria Rennes / IRISA and introduce them to the various facets of computer science. Camille Maumet and Pierre Maurel helped out to run the hands-on tutorials in which students could learn how a computer works without computers.

10.3.5. Other interventions

- Pierre-Yves Jonin: Series of 2 talks on neuropsychology, Universités du Temps Libre de Bretagne, Jan-Dec, Rennes, France.

10.3.6. Education

- L codent L créent - An outreach program to send PhD students to teach Python to middle school students in 8 sessions of 45 minutes. Camille Maumet is a co-organizer of the local version of this program, initiated in Lille. The program is currently supported by: Fondation Blaise Pascal, ED MathSTIC, Inria and Fondation Rennes 1.

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] O. COMMOWICK.*Compartments imaging for the characterization of brain diseases from quantitative MRI*, Université de Rennes 1, France, June 2019, Habilitation à diriger des recherches, https://www.hal.inserm.fr/tel-02173963

[2] C. MEURÉE.*Arterial spin labelling : quality control and super-resolution*, Université Rennes 1, March 2019, https://tel.archives-ouvertes.fr/tel-02281778

[3] H. SNOUSSI.*Imagerie de population et IRM de diffusion pour caractériser la sclérose en plaques pour la moélle épinière humaine*, Université Rennes 1, May 2019, https://tel.archives-ouvertes.fr/tel-02285896

Articles in International Peer-Reviewed Journal

[4] J. BEAUMONT, H. SAINT-JALMES, O. ACOSTA, T. KOBER, M. TANNER, J.-C. FERRÉ, O. SALVADO, J. FRIPP, G. GAMBAROTA.*Multi T1-weighted contrast MRI with fluid and white matter suppression at 1.5T*, in "Magnetic Resonance Imaging", November 2019, vol. 63, p. 217-225 [*DOI :* 10.1016/J.MRI.2019.08.010], https://hal-univ-rennes1.archives-ouvertes.fr/hal-02281467

[5] A.-C. BINTER, E. BANNIER, G. SIMON, D. SAINT-AMOUR, J.-C. FERRÉ, C. BARILLOT, C. MONFORT, S. CORDIER, C. CHEVRIER, F. PELÉ.*Prenatal exposure to glycol ethers and motor inhibition function evaluated by functional MRI at the age of 10 to 12years in the PELAGIE mother-child cohort*, in "Environment International", September 2019, vol. 133, n^{o} A, 105163 [*DOI :* 10.1016/J.ENVINT.2019.105163], https://www.hal.inserm.fr/inserm-02284944

[6] A. BOWRING, C. MAUMET, T. E. NICHOLS.*Exploring the Impact of Analysis Software on Task fMRI Results*, in "Human Brain Mapping", 2019, vol. 40, n^{o} 11, p. 3362-3384 [*DOI :* 10.1002/HBM.24603], https://www.hal.inserm.fr/inserm-01760535

[7] Y. CHAI, A. BUSH, J. COLOIGNER, A. NEDERVEEN, B. TAMRAZI, C. VU, S. CHOI, T. COATES, N. LEPORÉ, J. WOOD.*White matter has impaired resting oxygen delivery in sickle cell patients*, in "American Journal of Hematology", April 2019, vol. 94, n^{o} 4, p. 467-474 [*DOI :* 10.1002/AJH.25423], https://www.hal.inserm.fr/inserm-02174474

[8] R. Chouteau, B. Combès, E. Bannier, H. Snoussi, J.-C. Ferré, C. Barillot, G. Edan, P. Sauleau, A. Kerbrat.*Joint assessment of brain and spinal cord motor tract damage in patients with early RRMS: predominant impact of spinal cord lesions on motor function*, in "Journal of Neurology", June 2019, vol. 266, n° 9, p. 2294-2303 [*DOI :* 10.1007/s00415-019-09419-5], https://www.hal.inserm.fr/inserm-02160321

[9] J. Coloigner, J.-M. Batail, O. Commowick, I. Corouge, G. Robert, C. Barillot, D. Drapier.*White matter abnormalities in depression: a categorical and phenotypic diffusion MRI study*, in "Neuroimage-Clinical", February 2019, vol. 22, 101710 [*DOI :* 10.1016/j.nicl.2019.101710], https://www.hal.inserm.fr/inserm-01938407

[10] B. Combès, S. Prima.*An efficient EM-ICP algorithm for non-linear registration of large 3D point sets*, in "Computer Vision and Image Understanding", November 2019, p. 1-14 [*DOI :* 10.1016/j.cviu.2019.102854], https://hal.inria.fr/hal-02414127

[11] S. Crespo, M. Fasondini, C. Klein, N. Stoilov, C. Vallée.*Multidomain spectral method for the Gauss hypergeometric function*, in "Numerical Algorithms", June 2019, p. 1-35 [*DOI :* 10.1007/s11075-019-00741-7], https://hal-univ-bourgogne.archives-ouvertes.fr/hal-02194789

[12] C. Cury, S. Durrleman, D. Cash, M. Lorenzi, J. M. Nicholas, M. Bocchetta, J. C. Van Swieten, B. Borroni, D. Galimberti, M. Masellis, M. C. Tartaglia, J. Rowe, C. Graff, F. Tagliavini, G. B. Frisoni, R. Laforce, E. Finger, A. de Mendonça, S. Sorbi, S. Ourselin, J. Rohrer, M. Modat, C. Andersson, S. Archetti, A. Arighi, L. Benussi, S. Black, M. Cosseddu, M. Fallstrm, C. G. Ferreira, C. Fenoglio, N. Fox, M. Freedman, G. Fumagalli, S. Gazzina, R. Ghidoni, M. Grisoli, V. Jelic, L. Jiskoot, R. Keren, G. Lombardi, C. Maruta, L. Meeter, R. van Minkelen, B. Nacmias, L. Ijerstedt, A. Padovani, J. Panman, M. Pievani, C. Polito, E. Premi, S. Prioni, R. Rademakers, V. Redaelli, E. Rogaeva, G. Rossi, M. Rossor, E. Scarpini, D. Tang-Wai, H. Thonberg, P. Tiraboschi, A. Verdelho, J. Warren.*Spatiotemporal analysis for detection of pre-symptomatic shape changes in neurodegenerative diseases: Initial application to the GENFI cohort*, in "NeuroImage", March 2019, vol. 188, p. 282-290 [*DOI :* 10.1016/j.neuroimage.2018.11.063], https://www.hal.inserm.fr/inserm-01958916

[13] C. Cury, P. Maurel, R. Gribonval, C. Barillot.*A sparse EEG-informed fMRI model for hybrid EEG-fMRI neurofeedback prediction*, in "Frontiers in Neuroscience", 2020, forthcoming [*DOI :* 10.1101/599589], https://www.hal.inserm.fr/inserm-02090676

[14] D. Eden, C. Gros, A. Badji, S. Dupont, B. De Leener, J. Maranzano, R. Zhuoquiong, Y. Liu, T. Granberg, R. Ouellette, L. Stawiarz, J. Hillert, J. Talbott, E. Bannier, A. Kerbrat, G. Edan, P. Labauge, V. Callot, J. Pelletier, B. Audoin, H. Rasoanandrianina, J.-C. Brisset, P. Valsasina, M. Rocca, M. Filippi, R. Bakshi, S. Tauhid, F. Prados, M. Yiannakas, H. Kearney, O. Ciccarelli, S. Smith, C. Andrada Treaba, C. Mainero, J. Lefeuvre, D. Reich, G. Nair, T. Shepherd, E. Charlson, Y. Tachibana, M. Hori, K. Kamiya, L. Chougar, S. Narayanan, J. Cohen-Adad.*Spatial distribution of multiple sclerosis lesions in the cervical spinal cord*, in "Brain - A Journal of Neurology ", March 2019, vol. 142, n° 3, p. 633-646 [*DOI :* 10.1093/brain/awy352], https://www.hal.inserm.fr/inserm-02082374

[15] Y. Gautier, P. Meurice, N. Coquery, A. Constant, E. Bannier, Y. Serrand, J.-C. Ferré, R. Moirand, D. Val-Laillet.*Implementation of a New Food Picture Database in the Context of fMRI and Visual Cognitive Food-Choice Task in Healthy Volunteers*, in "Frontiers in Psychology",

November 2019, vol. 10, 2660, Yentl Gautier and Paul Meurice have contributed equally to this work [*DOI :* 10.3389/FPSYG.2019.02620], https://www.hal.inserm.fr/inserm-02389926

[16] C. MEURÉE, P. MAUREL, J.-C. FERRÉ, C. BARILLOT.*Patch-Based Super-Resolution of Arterial Spin Labeling Magnetic Resonance Images*, in "NeuroImage", January 2019, vol. 189, p. 85-94 [*DOI :* 10.1016/J.NEUROIMAGE.2019.01.004], https://www.hal.inserm.fr/inserm-01880726

[17] L. PIEROT, C. BARBE, J.-C. FERRÉ, C. COGNARD, S. SOIZE, P. WHITE, L. SPELLE.*Patient and aneurysm factors associated with aneurysm rupture in the population of the ARETA study*, in "American Journal of Neuroradiology", September 2019 [*DOI :* 10.1016/J.NEURAD.2019.07.007], https://hal-univ-rennes1.archives-ouvertes.fr/hal-02371090

[18] M. PROISY, I. COROUGE, A. LEGOUHY, A. NICOLAS, V. CHARON, N. MAZILLE, S. LEROUX, B. BRUNEAU, C. BARILLOT, J.-C. FERRÉ.*Changes in brain perfusion in successive arterial spin labeling MRI scans in neonates with hypoxic-ischemic encephalopathy*, in "Neuroimage-Clinical", July 2019, vol. 101939, p. 1-31 [*DOI :* 10.1016/J.NICL.2019.101939], https://www.hal.inserm.fr/inserm-02189386

Invited Conferences

[19] E. BANNIER, B. COMBÈS, A. KERBRAT.*Multi-center (mono-vendor) longitudinal conventional and quantitative spinal cord MRI in Multiple Sclerosis at 3 Tesla - The EMISEP Study : First results*, in "2019 - 6th Spinal Cord MRI Workshop", Londres, United Kingdom, January 2019, p. 1-30, https://www.hal.inserm.fr/inserm-01994583

[20] C. MAUMET.*Let's do open neuroimaging sciences*, in "GEANT 2019 - Workshop GEstion et pArtage de données en NeuroinformaTique", Marseille, France, May 2019, p. 1-54, https://www.hal.inserm.fr/inserm-02146290

International Conferences with Proceedings

[21] S. BUTET, G. LIOI, M. FLEURY, A. LÉCUYER, C. BARILLOT, I. BONAN.*A multi-target motor imagery training using EEG-fMRI Neurofeedback: an exploratory study on stroke*, in "OHBM 2019- Orgaization for Human Brain Mapping", Rome, Italy, June 2019, p. 1-4, https://hal.inria.fr/hal-02265496

[22] A. LEGOUHY, O. COMMOWICK, F. ROUSSEAU, C. BARILLOT.*Unbiased longitudinal brain atlas creation using robust linear registration and log-Euclidean framework for diffeomorphisms*, in "ISBI 2019 - 16th IEEE International Symposium on Biomedical Imaging", Venise, Italy, IEEE International Symposium on Biomedical Imaging (ISBI), IEEE, April 2019, p. 1038-1041 [*DOI :* 10.1109/ISBI.2019.8759508], https://www.hal.inserm.fr/inserm-02099958

[23] G. LIOI, S. BUTET, M. FLEURY, C. CURY, E. BANNIER, A. LÉCUYER, I. BONAN, C. BARILLOT.*Bimodal EEG-fMRI Neurofeedback for upper motor limb rehabilitation: a pilot study on chronic patients*, in "rtFIN 2019 - Real Time Functional Imaging and Neurofeedback", Maastricht, Netherlands, December 2019, p. 1-2, https://hal.inria.fr/hal-02383532

Conferences without Proceedings

[24] A. BATES, A. DADUCCI, E. CARUYER.*Multi-Dimensional Diffusion MRI Sampling Scheme: B-tensor Design and Accurate Signal Reconstruction*, in "ISMRM 2019 - 27th Annual Meeting & Exhibition", Montréal, Canada, May 2019, p. 1-4, https://www.hal.inserm.fr/inserm-02065830

[25] J. BEAUMONT, H. SAINT-JALMES, O. ACOSTA, T. KOBER, M. TANNER, J.-C. FERRÉ, O. SALVADO, J. FRIPP, G. GAMBAROTA.*High Contrast T1-Weigthed Mri with Fluid and White Matter Suppression Using Mp2Rage*, in "ISBI 2019 - 16th IEEE International Symposium on Biomedical Imaging", Venice, Italy, IEEE, 2019, p. 701-704 [*DOI :* 10.1109/ISBI.2019.8759494], https://hal-univ-rennes1.archives-ouvertes.fr/hal-02391261

[26] S. CHATTERJEE, O. COMMOWICK, O. AFACAN, S. K. WARFIELD, C. BARILLOT.*Tissue microstructure information from T2 relaxometry and diffusion MRI can identify multiple sclerosis (MS) lesions undergoing blood-brain barrier breakdown (BBB)*, in "ISMRM 2019 - 27th Annual Meeting & Exhibition", Montreal, Canada, May 2019, p. 1-4, https://www.hal.inserm.fr/inserm-02153311

[27] J. COLOIGNER, J.-M. BATAIL, D. DRAPIER, C. BARILLOT.*Structural connectivity analysis in treatment-resistant depression*, in "OHBM 2019 - 25th Annual Meeting of the Organization for Human Brain Mapping", Rome, Italy, June 2019, p. 1-2, https://hal-univ-rennes1.archives-ouvertes.fr/hal-02176403

[28] J. COLOIGNER, A.-C. BINTER, E. BANNIER, J.-C. FERRÉ, C. CHEVRIER, C. BARILLOT, F. PELÉ.*Effect of prenatal organic solvent exposure on structural connectivity at childhood*, in "ISBI 2019 - 16th IEEE International Symposium on Biomedical Imaging", Venise, Italy, IEEE, April 2019, p. 151-154 [*DOI :* 10.1109/ISBI.2019.8759473], https://www.hal.inserm.fr/inserm-02176376

[29] C. CURY, P. MAUREL, R. GRIBONVAL, C. BARILLOT.*Can we learn from coupling EEG-fMRI to enhance neuro-feedback in EEG only?*, in "OHBM 2019 - Annual Meeting Organization for Human Brain Mapping", Rome, Italy, June 2019, 1, https://www.hal.inserm.fr/inserm-02074623

[30] C. CURY, P. MAUREL, G. LIOI, R. GRIBONVAL, C. BARILLOT.*Learning bi-modal EEG-fMRI neurofeedback to improve neurofeedback in EEG only*, in "Real-Time Functional Imaging and Neurofeedback", Maastricht, Netherlands, December 2019, p. 1-2 [*DOI :* 10.1101/599589], https://www.hal.inserm.fr/inserm-02368720

[31] C. CURY, M. A. SCELZI, R. TORO, V. FROUIN, E. ARTIGES, A. HEINZ, H. W. WALTER, H. LEMAÎTRE, J.-L. MARTINOT, J.-B. POLINE, M. N. SMOLKA, G. SCHUMANN, A. ALTMANN, O. COLLIOT.*On the genetic bases of incomplete hippocampal inversion: a genome-wide association study*, in "OHBM 2019 - Annual Meeting on Organization for Human Brain Mapping", Rome, Italy, June 2019, 1, https://www.hal.inserm.fr/inserm-02074616

[32] Q. DUCHÉ, F. CHAPELAIN, E. BANNIER, J.-C. FERRÉ, P. ALLAIN, P. GALLIEN, V. DARDIER, C. BARILLOT.*Neural basis of irony in patients with Multiple Sclerosis : an exploratory fMRI study*, in "OHBM 2019 - 25th Annual Meeting of the Organization for Human Brain Mapping", Rome, Italy, June 2019, p. 1-3, https://www.hal.inserm.fr/inserm-02148314

[33] M. FLEURY, G. LIOI, C. BARILLOT, A. LÉCUYER.*The use of haptic feedback in Brain-Computer Interfaces and Neurofeedback*, in "rtFIN 2019 - Real Time Functional Imaging and Neurofeedback", Maastricht, Netherlands, December 2019, https://hal.archives-ouvertes.fr/hal-02387400

[34] M. FLEURY, G. LIOI, C. BARILLOT, A. LÉCUYER.*The use of haptic feedback in Brain-Computer Interfaces and Neurofeedback*, in "CORTICO 2019 - COllectif pour la Recherche Transdisciplinaire sur les Interfaces Cerveau-Ordinateur", Lille, France, March 2019, https://hal.archives-ouvertes.fr/hal-02387408

[35] F. GALASSI, S. TARRIDE, E. VALLÉE, O. COMMOWICK, C. BARILLOT.*Deep learning for multi-site ms lesions segmentation: two-step intensity standardization and generalized loss function*, in "ISBI 2019 - 16th

IEEE International Symposium on Biomedical Imaging", Venice, Italy, IEEE, April 2019, 1, https://hal.archives-ouvertes.fr/hal-02052250

[36] K. HELMER, D. KEATOR, T. AUER, S. GHOSH, C. MAUMET, T. NICHOLS, J.-B. POLINE.*Constructing an Ontology of Neuroscience Experiments for the Neuroimaging Data Model (NIDM) Authors: Introduction*, in "OHBM 2019 - 25th Annual Meeting of the Organization for Human Brain Mapping, Jun 2019", Rome, Italy, June 2019, p. 1-4, https://www.hal.inserm.fr/inserm-02379281

[37] P.-Y. JONIN, Q. DUCHÉ, E. BANNIER, I. COROUGE, J.-C. FERRÉ, S. BELLIARD, E. J. BARBEAU, C. BARILLOT.*Abnormal fMRI response in sub-hippocampal structures: how prior knowledge impairs memory in AD*, in "OHBM 2019 - 25th Annual Meeting of the Organization for Human Brain Mapping", Rome, Italy, June 2019, p. 1-5, https://www.hal.inserm.fr/inserm-02148289

[38] P.-Y. JONIN, Q. DUCHÉ, E. BANNIER, I. COROUGE, J.-C. FERRÉ, S. BELLIARD, C. BARILLOT, E. J. BARBEAU.*Learning What You Know: How Prior Knowledge Impairs New Associative Learning in Early AD*, in "2019 Conference on Cognitive Neurosciences of Memory: Recollection, Familiarity and Novelty detection", Liège, Belgium, October 2019, 1, https://www.hal.inserm.fr/inserm-02383947

[39] P.-Y. JONIN, Q. DUCHÉ, E. BANNIER, I. COROUGE, J.-C. FERRÉ, S. BELLIARD, C. BARILLOT, E. J. BARBEAU.*Learning What You Know: How Prior Knowledge Impairs New Associative Learning in Early AD*, in "AAIC 2019 - Alzheimer's Association International Conference", Los Angeles, United States, July 2019, p. 1-2, https://www.hal.inserm.fr/inserm-02383941

[40] D. KEATOR, K. HELMER, C. MAUMET, S. PADHY, D. JARECKA, D. N. KENNEDY, S. GHOSH, J.-B. POLINE.*Tools for FAIR Neuroimaging Experiment Metadata Annotation with NIDM Experiment*, in "OHBM 2019 - 25th Annual Meeting of the Organization for Human Brain Mapping,", Rome, Italy, June 2019, p. 1-5, https://www.hal.inserm.fr/inserm-02379309

[41] S. LE FRANC, M. FLEURY, M. COGNÉ, S. BUTET, C. BARILLOT, A. LÉCUYER, I. BONAN.*Influence of visual feedback on the illusion of movement induced by tendon vibration of wrist in healthy subjects*, in "SOFMER 2019 - 34ème congrès de la Société Français de Médecine Physique et de Réadaptation", Bordeaux, France, October 2019, https://hal.inria.fr/hal-02415992

[42] A. LEGOUHY, O. COMMOWICK, F. ROUSSEAU, C. BARILLOT.*Online atlasing using an iterative centroid*, in "MICCAI 2019 - 22nd International Conference on Medical Image Computing and Computer Assisted Intervention", Shenzhen, China, October 2019, p. 366-374 [*DOI :* 10.1007/978-3-030-32248-9_41], https://www.hal.inserm.fr/inserm-02317392

[43] G. LIOI, S. BUTET, M. FLEURY, A. LÉCUYER, I. BONAN, C. BARILLOT.*Efficacy of EEG-fMRI Neurofeedback in stroke in relation to the DTI structural damage: a pilot study*, in "OHBM 2019 - 25th Annual Meeting of the Organization for Human Brain Mapping", Rome, Italy, June 2019, p. 1-4, https://hal.inria.fr/hal-02265495

[44] C. MAUMET, S. GHOSH, Y. O. HALCHENKO, D. JARECKA, N. B. NICHOLS, J.-B. POLINE, M. HANKE.*The best of both worlds: using semantic web with JSON-LD. An example with NIDM-Results & Datalad*, in "OHBM 2019 - 25th Annual Meeting of the Organization for Human Brain Mapping", Rome, Italy, June 2019, https://www.hal.inserm.fr/inserm-01972649

[45] C. MAUMET, X. ROLLAND, A. BONNET, S. CAMARASU-POP, A. KOKOGIANNAKI, C. BARILLOT.*Open research: linking the bits and pieces with OpenAIRE-connect*, in "OHBM 2019 - 25th Annual Meeting of the Organization for Human Brain Mapping", Rome, Italy, June 2019, p. 1-6, https://www.hal.inserm.fr/inserm-02151177

[46] T. POSTMA, C. CURY, S. BAXENDALE, J. P. THOMPSON, I. CANO-LOPEZ, J. DE TISI, J. L. BURDETT, M. K. SIDHU, L. CACIAGLI, G. P. WINSTON, S. B. VOS, M. THOM, J. S. DUNCAN, M. J. KOEPP, M. GALOVIC.*Hippocampal Shape Alterations Are Associated With Pre- And Postsurgical Verbal And Visual Memory Deficits In Temporal Lobe Epilepsy*, in "IEC 2019 - 33rd International Epilepsy Congress", Bangkok, Thailand, June 2019, 1, https://hal.archives-ouvertes.fr/hal-02383692

[47] M. PROISY, I. COROUGE, A. LEGOUHY, V. CHARON, N. MAZILLE, A. NICOLAS, S. LEROUX, B. BRUNEAU, C. BARILLOT, J.-C. FERRÉ.*Changes in brain perfusion in successive arterial spin labelling MRI scans in neonates with hypoxic-ischemic encephalopathy*, in "SPR 2019 - Annual Meeting & Postgraduate Course", San Francisco, California, United States, April 2019, https://www.hal.inserm.fr/inserm-01944507

[48] H. SNOUSSI, E. CARUYER, J. COHEN-ADAD, O. COMMOWICK, B. COMBÈS, E. BANNIER, A. KERBRAT, C. BARILLOT.*Geometric evaluation of distortion correction methods in diffusion MRI of the spinal cord*, in "ISBI 2019 - 16th IEEE International Symposium on Biomedical Imaging", Venice, Italy, IEEE, April 2019, p. 1696-1699 [*DOI :* 10.1109/ISBI.2019.8759196], https://www.hal.inserm.fr/inserm-01986723

[49] R. TRUFFET, C. BARILLOT, E. CARUYER.*Optimal selection of diffusion-weighting gradient waveforms using compressed sensing and dictionary learning*, in "ISMRM 2019 - 27th Annual Meeting & Exhibition", Montréal, Canada, May 2019, p. 1-3, https://www.hal.inserm.fr/inserm-02015394

[50] C. VALLÉE, P. MAUREL, I. COROUGE, C. BARILLOT.*Investigating an optimal duration of acquisition in resting-state ASL*, in "OHBM 2019 - 25th Annual Meeting of the Organization for Human Brain Mapping", Rome, Italy, June 2019, p. 1-3, https://www.hal.inserm.fr/inserm-02150255

Scientific Popularization

[51] C. BARILLOT, V. EMILIANI, M. TANTER.*Imagerie et analyses fonctionnelles*, in "Le Cerveau en lumières", B. POULAIN, E. HIRSCH (editors), Odile Jacob, 2019, https://www.hal.inserm.fr/inserm-02414949

[52] C. CURY.*The hippocampus in all its forms, travel in the heart of the brain*, March 2019, p. 1-33, Semaine du Cerveau 2019 - brain Awareness Week, https://hal.inria.fr/hal-02075039

[53] C. MAUMET.*Cartographes du cerveau*, March 2019, p. 1-44, Semaine du Cerveau 2019 - brain Awareness Week, https://hal.inria.fr/hal-02078743

Other Publications

[54] A. ACKAOUY, N. COURTY, E. VALLEE, O. COMMOWICK, C. BARILLOT, F. GALASSI.*Unsupervised Domain Adaptation with Optimal Transport in multi-site segmentation of Multiple Sclerosis lesions from MRI data : Preprint*, October 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02317028

[55] G. BESSON, P.-Y. JONIN, C. BARILLOT.*Wakeful rest promotes the viewpoint invariance, but not the fine discriminability, of entities mnemonic representations for subsequently familiarity-based recognition*, October 2019, 2019 - Cognitive Neuroscience of Memory: Recollection Familiarity Novelty detection, Poster, https://www.hal.inserm.fr/inserm-02384508

[56] A. GUYOT, A. B. GRACIANO FOUQUIER, E. GERARDIN, M. CHUPIN, J. GLAUNÈS, L. MARRAKCHI-KACEM, J. GERMAIN, C. BOUTET, C. CURY, L. HERTZ-PANNIER, A. VIGNAUD, S. DURRLEMAN, T. HENRY, P.-F. VAN DE MOORTELE, A. TROUVÉ, O. COLLIOT.*A Diffeomorphic Vector Field Approach to Analyze the Thickness of the Hippocampus from 7T MRI*, November 2019, working paper or preprint, https://hal.inria.fr/hal-02359660

[57] A. A. OULD ISMAIL, D. PARKER, M. HERNANDEZ-FERNANDEZ, R. WOLF, S. BREM, S. ALEXANDER, W. HODGES, O. PASTERNAK, E. CARUYER, R. VERMA.*Freewater EstimatoR using iNtErpolated iniTialization (FERNET): Toward Accurate Estimation of Free Water in Peritumoral Region Using Single-Shell Diffusion MRI Data*, October 2019, working paper or preprint [*DOI :* 10.1101/796615], https://www.hal.inserm.fr/inserm-02309753

[58] C. VALLÉE, P. MAUREL, I. COROUGE, C. BARILLOT.*Acquisition duration in resting-state arterial spin labeling. How long is enough?*, September 2019, working paper or preprint, https://www.hal.inserm.fr/inserm-02129799

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

Institut national de recherche pour l'agriculture, l'alimentation et l'environnement

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Earth, Environmental and Energy Sciences

Table of contents

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. - Methods in mathematical modeling
A6.1.2. - Stochastic Modeling
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. Team, Visitors, External Collaborators

Research Scientists

Etienne Mémin [Team leader, Inria, Senior Researcher, HDR]
Jocelyne Chaux [Inria, Senior Researcher, HDR]
Christophe Collewet [IRSTEA, Researcher, HDR]
Dominique Heitz [IRSTEA, Researcher, HDR]
Gilles Tissot [Inria, Researcher]

Faculty Member

Roger Lewandowski [Univ de Rennes I, Professor, HDR]

Technical Staff

Werner Bauer [Inria, Engineer, until Oct 2019]

Pranav Chandramouli [Inria, Engineer, from Feb 2019]

PhD Students

Mohamed Yacine Ben Ali [CSTB, PhD Student]
Bastien Hamlat [IFPEN, PhD Student, until Sep 2019]
Long Li [Univ de Rennes I, PhD Student]
Romain Schuster [ITGA, PhD Student, until Jan 2019, granted by CIFRE]

Post-Doctoral Fellow

Carlo Cintolesi [Inria, Post-Doctoral Fellow, until Feb 2019]

Visiting Scientist

Rudiger Brecht [Inria, from May 2019 until Aug 2019]

Administrative Assistant

Huguette Bechu [Inria, Administrative Assistant]

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 [56] 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 [67]. 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, p ollutant, light index refraction, fluorescein,...). 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 [61]. Dynamically consistent and highly robust techniques have been also proposed for the recovery of surface oceanic streams from satellite images [58]. Very recently parameter-free approaches relying on uncertainty concept has been devised [59]. 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 [68], [69] or through stochastic filtering approaches [62], [65]. 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 [60] 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 [72], [64]. 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 [57]. Most of the works of the literature indeed comes back to open-loop control approaches [71], [66], [70] or to forcing approaches [63] 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. Highlights of the Year

4.1. Highlights of the Year

4.1.1. Awards

Best paper award 2019 Romain Schuster "Visualisation et mesure du flux d'aspiration d'une Sorbonne", ContaminExpert 2019. Paris, FR
BEST PAPERS AWARDS :

[43]
R. SCHUSTER, D. HEITZ, E. MÉMIN.*Visualisation et mesure du flux d'aspiration d'une Sorbonne*, in "ContaminExpert 2019", Paris, France, March 2019, p. 1-8, https://hal.archives-ouvertes.fr/hal-02330348

5. New Software and Platforms

5.1. 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

5.2. 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

5.3. 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

5.4. 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

5.5. TYPHOON

KEYWORD: Fluid mechanics

FUNCTIONAL DESCRIPTION: Typhoon is a fluid motion estimator from image sequences. It is almost real-time dedicated to the measurement of LIDAR sequences, multi-scale, fast and precise to make a fine scale analysis of fluid flows with applications in the fields of energy, transport and environment.

- Participants: Christopher Mauzey, Étienne Mémin and Pierre Dérian
- Partner: CSU Chico
- Contact: Étienne Mémin
- URL: https://phys.csuchico.edu/lidar/typhoon/

5.6. H2OLab

KEYWORDS: Simulation - Energy - Contamination - Groundwater - Hydrogeology - Heterogeneity - Uncertainly - Multiscale

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/

5.7. 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
- Contact: Géraldine Pichot

5.8. GRT3D

Global Reactive Transport in 3D

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

6. New Results

6.1. Fluid motion estimation

6.1.1. Stochastic uncertainty models for motion estimation

Participants: Musaab Khalid Osman Mohammed, Etienne Mémin.

This work is concerned with the design of motion estimation technique for image-based river velocimetry. The method proposed is based on an advection diffusion equation associated to the transport of large-scale quantity with a model of the unresolved small-scale contributions. Additionally, since there is no ground truth data for such type of image sequences, a new evaluation method to assess the results has been developed. It is based on trajectory reconstruction of few Lagrangian particles of interest and a direct comparison against their manually-reconstructed trajectories. The new motion estimation technique outperformed traditional optical flow and PIV-based methods used in hydrology [23]. This study has been performed within the PhD thesis of Musaab Khalid and through a collaboration with the Irstea Lyon hydrology research group (HHLY).

6.1.2. 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. The technique was validated on a mixing layer in a wind tunnel for HFSB and smoke tracers [39] and applied on a laboratory fume-hood [26], [30], [43]. This study demonstrated the feasibility of conducting on-site large-scale image-based measurements for indoor airflows characterization. The technique was also assessed in an outdoor flow

6.1.3. 3D flows reconstruction from image data

Participants: Dominique Heitz, Etienne Mémin.

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 via ensemble assimilation scheme. We have proposed a novel method for volumetric velocity reconstruction exploring the locality of 3D object space.

Under this formulation the velocity of local patch was sought to match the projection of the particles within the local patch in image space to the image recorded by camera. The core algorithm to solve the matching problem is an instance-based estimation scheme that can overcome the difficulties of optimization originated from the nonlinear relationship between the imageintensity residual and the volumetric velocity. The proposed method labeled as Lagrangian Particle ImageVelocimetry (LaPIV) is quantitatively evaluated with synthetic particle image data. The promising results indicated the potential application of LaPIV to a large variety of volumetric velocity reconstruction problems [44].

6.2. Tracking, Data assimilation and model-data coupling

6.2.1. 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, Gilles Tissot.

In this axis of work, we explore 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 planes 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 [21]. We extended 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 is 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.

6.2.2. Ensemble variational data assimilation of large-scale dynamics with uncertainty

Participant: Etienne Mémin.

Estimating the parameters of geophysical dynamic models is an important task in Data Assimilation (DA) technique used for forecast initialization and reanalysis. In the past, most parameter estimation strategies were derived by state augmentation, yielding algorithms that are easy to implement but may exhibit convergence difficulties. The Expectation-Maximization (EM) algorithm is considered advantageous because it employs two iterative steps to estimate the model state and the model parameter separately. In this work, we propose a novel ensemble formulation of the Maximization step in EM that allows a direct optimal estimation of physical parameters using iterative methods for linear systems. This departs from current EM formulations that are only capable of dealing with additive model error structures. This contribution shows how the EM technique can be used for dynamics identification problem with a model error parameterized as arbitrary complex form. The proposed technique is used for the identification of stochastic subgrid terms that account for processes unresolved by a geophysical fluid model. This method, along with the augmented state technique, has been evaluated to estimate such subgrid terms through high resolution data. Compared to the augmented state technique, our method is shown to yield considerably more accurate parameters. In addition, in terms of prediction capacity, it leads to smaller generalization error as caused by the overfitting of the trained model on presented data and eventually better forecasts [29].

6.2.3. Reduced-order models for flows representation from image data

Participants: Dominique Heitz, Etienne Mémin, Gilles Tissot.

During the PhD thesis of Valentin Resseguier we have 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 as 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. 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 [15]. Note that these reduced order models can be extended to a full system of stochastic differential equations 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 as meaningful physical behavior with respect to the simulation of the neutral modes of the dynamics. This study is pursued within a strong collaboration with the industrial partner: SCALIAN

6.2.4. Learning of the dynamics of large scale geophysical systems using semi-group theory for data assimilation

Participants: Etienne Mémin, Gilles Tissot.

The goal of this study is to propose new ensemble data assimilation methodologies to estimate oceanic and turbulent flows. In classical methods, from a distribution of initial conditions, an ensemble of simulations are computed and used for estimation. Ideally, from this solution, a new ensemble has to be generated to refine the estimation. However, due to large numerical costs and operational constraints, this iterative procedure is in practice intractable. In order to improve actual performances, we propose to take these limitations into account and to develop new methodologies able to better take advantage of the information contained in the ensemble and in the dynamical model. More precisely, we propose to learn the non-linear dynamical features of the system and to be able to reproduce it without having to run a new simulation. The formalism is based on two concepts: i) the reproducing kernel Hilbert spaces (RKHS) that are a basis of smooth functions in the phase space giving interpolatory properties ii) the Koopman operator, that is an infinite-dimensional operator able to propagate in time any observable of the phase space. These two elements allow to define a rigorous framework in which hypothesis classically done in ensemble methods appear naturally. Thus, classical methods enter in a special case of this new formalism, that allows us to generalise them in a way to improve the learning of the non-linear dynamical system. Numerical tests are performed using the Ginzburg-Landau equation and a quasi-geostrophic flow model.

6.2.5. Estimation and control of amplifier flows

Participant: Gilles Tissot.

Estimation and control of fluid systems is an extremely hard problem. The use of models in combination with data is central to take advantage of all information we have on the system. Unfortunately all flows do not present the same physical and mathematical behaviour, thus using models and methodologies specialised to the flow physics is necessary to reach high performances.

A class of flows, denoted "oscillator flows", are characterised by unstable modes of the linearised operator. A consequence is the dominance of relatively regular oscillations associated with a nonlinear saturation. Despite the non-linear behaviour, associated structures and dynamical evolution are relatively easy to predict. Canonical configurations are the cylinder wake flow or the flow over an open cavity.

By opposition to that, "amplifier flows" are linearly stable with regard to the linearised operator. However, due to their convective nature, a wide range of perturbations are amplified in time and convected away such that it vanishes at long time. The consequence is the high sensitivity to perturbations and the broad band response that forbid a low rank representation. Jets and mixing layers show this behaviour and a wide range of industrial applications are affected by these broad band perturbations. It constitutes then a class of problems that are worth to treat separately since it is one of the scientific locks that render hard the transfer of methodologies existing in flow control and estimation to industrial applications.

There exists a type of models, that we will denote as "parabolised", that are able to efficiently represent amplifier flows. These models, such as parabolised stability equations and one-way Navier-Stokes propagate, in the frequency domain, hydrodynamic instability waves over a given turbulent mean flow. We can note that these models, by their structure, give access to a natural experimental implementation. They are an ingredient adapted to represent the system, but have a mathematical structure strongly different from the dynamical models classically used in control and data assimilation. It is then important to develop new methodologies of control, estimation and data assimilation with these models to reach our objectives. Moreover, inventing new models by introducing the modelling under location uncertainties in these parabolised models will be perfectly adapted to represent the evolution and the variability of an instability propagating within a turbulent flow. It will be consistent with actual postprocessing of experimental data performed in similar flow configurations.

6.3. Analysis and modeling of turbulent flows and geophysical flows

6.3.1. Geophysical flows modeling under location uncertainty

Participants: Werner Bauer, Pranav Chandramouli, Long Li, Etienne Mémin.

In this research axis we have devised a principle to derive representation of flow dynamics under location 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. Rigorously derived from a stochastic version of the Reynolds transport theorem [9], this framework, referred to as modeling under location uncertainty (LU), 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 LU modeling can be applied to provide stochastic representations of a variety of classical geophysical flows dynamics [12], [13], [14]. Numerical simulations and uncertainty quantification have been performed on Quasi Geostophic approximation (QG) of oceanic models. It has been shown that LU leads to remarkable estimation of the unresolved errors opposite 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). For a wind forced QG model in a square box, which is an idealized model of north-Atlantic circulation, we have shown that for different versions of the noise the QG LU model leads to improve long-terms statistics when compared to classical large-eddies simulation strategies. For a QG model we have demonstrated that the LU model allows conserving the global energy. We have also shown numerically that Rosby waves were conserved and that inhomogeneity of the random component triggers secondary circulations. This feature enabled us to draw a formal bridge between a classical system describing the interactions between the mean current and the surface waves and the LU model in which the turbophoresis advection term plays the role of the classical Stokes drift.

Supported by funding from Inria-Mitacs Globalink, we hosted Ruediger Brecht, PhD student at Memorial University of Newfoundland, Canada, for a period of 3 months (May to August) in the Fluminance group. During his stay, Ruediger Brecht worked on the incorporation of a stochastic representation of the small-scale velocity component of a fluid flow in a variational integrator for the rotating shallow-water equations on the sphere, already developed within the first part of its PhD work. This work was based on an ongoing study in the group on a stochastic Quasi-geostrophic model and followed a series of works performed in the Fluminance group to define stochastic geophysical flow dynamics.

6.3.2. Large eddies simulation models under location uncertainty

Participants: Mohamed Yacine Ben Ali, Pranav Chandramouli, Dominique Heitz, Etienne Mémin, Gilles Tissot.

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 a flow over a circular cylinder at Re 3900 and wall-bounded flows. 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.

6.3.3. Variational principles for structure-preserving discretizations in stochastic fluid dynamics

Participants: Werner Bauer, Long Li, Etienne Mémin.

The overarching goal of this interdisciplinary project is to use variational principles to derive deterministic and stochastic models and corresponding accurate and efficient structure preserving discretizations and to use these schemes to obtain a deeper understanding of the conservation laws of the stochastic fluid dynamics investigated. The newly developed systematic discretization framework is based on discrete variational principles whose highly structured procedures shall be exploited to develop a general software framework that applies automatic code generation. This project will first provide new stochastic fluid models and suitable approximations, with potential future applications in climate science using the developed methods to perform accurate long term simulations while quantifying the solutions uncertainties. The generality of our approach addresses also other research areas such as electrodynamics (EDyn), magnetohydrodynamics (MHD), and plasma physics.

6.3.4. Stochastic compressible fluid dynamics

Participants: Etienne Mémin, Gilles Tissot.

Some work has been performed to extend the stochastic formulation under location uncertainty to compressible flows. The interest is to extend the formulation on the one hand to compressible fluids (for instability mechanisms involved in areoacoustics for instance, or for thermal effects in mixing layers) and on the other hand to geophysical flows where the Boussinesq equation is not valid anymore (density variations due to temperature or salinity gradients). A theoretical study has been performed that opens the door to numerical validations. In particular a baroclinic torque term has been identified that could have major effects in some situations.

6.3.5. Stochastic hydrodynamic stability under location uncertainty

Participants: Etienne Mémin, Gilles Tissot.

In order to predict instability waves propagating within turbulent flows, eigenmodes of the linearised operator is not well suited since it neglects the effect of turbulent fluctuations on the wave dynamics. To cope this difficulty, resolvent analysis has become popular since it represents the response of the linearised operator to any forcing representing the generalised stress tensors. The absence of information on the non-linearity is a strong limitation of the method. In order to refine these models, we propose to consider a stochastic model under location uncertainty expressed in the Fourier domain, to linearise it around the corrected mean-flow and to study resulting eigenmodes. The stochastic part represents the effect of the turbulent field onto the instability wave. It allows to specify a structure of the noise and then to improve existing models. Improvements compared to the resolvent analysis have been found for turbulent channel flow data at $\mathfrak{R}_\tau = 180$. This work is in collaboration with André Cavalieri (Instituto Tecnologico de Aeronautica, SP, Brésil).

6.3.6. 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? 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,carried when the regularizing parameter goes to 0. These results are described in a paper published in JMAA [24]

In the framework of the collaboration with the University of Pisa (Italy), namely with Luigi Berselli collaboration, we considered the three dimensional incompressible Navier-Stokes equations with non stationary source terms chosen in a suitable space. We proved the existence of Leray-Hopf weak solutions and that it is possible to characterize (up to sub-sequences) their long-time averages, which satisfy the Reynolds averaged equations, involving a Reynolds stress. Moreover, we showed that the turbulent dissipation is bounded by the sum of the Reynolds stress work and of the external turbulent fluxes, without any additional assumption, than that of dealing with Leray-Hopf weak solutions. This is a very nice generalisation to non stationnary source terms of a famous results by Foais. IN the same work, we also considered ensemble averages of solutions, associated with a set of different forces and we proved that the fluctuations continue to have a dissipative effect on the mean flow. These results have been published in Nonlinearity [19]. These results have been extended in the framework of POD for reduced models in [18].

In [55] we have shown the existence of a solution to a 1D Reynolds Averaged Navier-Stokes vertical model suitable in the atmospheric boundary layer, under suyitable assumption on the data. The paper is received for publication in thje journal Pure and Applied Functional Analysis (PAFA).

We also have introduced a turbulence model including a backscatter term, which has the same structure as the Voigt model. The additional term is derived in certain specific regimes of the flow, such as the convergence to stable statistical states. We get estimates for the velocity v in $L_t^\infty H_x^1 \cap W_t^{1,2} H_x^{1/2}$, that allow us to prove the existence and uniqueness of a regular-weak solutions (v, p) to the resulting system, for a given fixed eddy viscosity. We then prove a structural compactness result that highlights the robustness of the model. This allows us to pass to the limit in the quadratic source term in the equation for the turbulent kinetic energy k, which yields the existence of a weak solution to the corresponding Reynolds Averaged Navier-Stokes system satisfied by (v, p, k). These results are writen in [47], a paper which is under revision in Non Linear Analysis.

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 have tested the performances of an incompressible turbulence Reynolds-Averaged Navier-Stokes one-closure equation model in a boundary layer, which requires the determination of the mixing length l. A series of direct numerical simulation have been performed, with flat and non trivial topographies, to obtain by interpolation a generic formula l = l(Re6, z), Re6 being the frictional Reynolds number, and z the distance to the wall. Numerical simulations have been carried out at high Reynolds numbers with this turbulence model, in order to discuss its ability to properly reproduce the standard profiles observed in neutral boundary layers, and to assess its advantages, its disadvantages and its limits. We also proceeded to a mathematical analysis of the model.

6.3.7. Stochastic flow model to predict the mean velocity in wall bounded flows

Participants: Roger Lewandowski, Etienne Mémin, Benoit Pinier.

To date no satisfying model exists to explain the mean velocity profile within the whole turbulent layer of canonical wall bounded flows. We propose a modification of the velocity profile expression that ensues from the stochastic representation of fluid flows dynamics proposed recently in the group and referred to as "modeling under location uncertainty". This framework introduces in a rigorous way a subgrid term generalizing the eddy-viscosity assumption and an eddy-induced advection term resulting from turbulence inhomogeneity. This latter term gives rise to a theoretically well-grounded model for the transitional zone between the viscous sublayer and the turbulent sublayer. An expression of the small-scale velocity component is also provided in the viscous zone. Numerical assessment of the results have been performed for turbulent boundary layer flows, pipe flows and channel flows at various Reynolds numbers [25][17].

6.3.8. 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 three-dimensional 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 carry 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 have been be used within a 4DVar assimilation technique, to reconstruct three-dimensional turbulent flows from data. The third challenge was to carry out a time resolved tomoPIV experiments in a turbulent wake flow. This work has been submitted to the Journal of Computational Physics.

6.3.9. Fast 3D flow reconstruction from 2D cross-plane observations

Participants: Pranav Chandramouli, Dominique Heitz, Etienne Mémin.

We proposed a computationally efficient flow reconstruction technique, exploiting homogeneity in a given direction, to recreate three dimensional instantaneous turbulent velocity fields from snapshots of two dimension planar fields. This methodology, termed as "snapshot optimisation" or SO, enables to provide 3D data-sets for studies which are currently restricted by the limitations of experimental measurement techniques. The SO method aims at optimising the error between an inlet plane with a homogeneous direction and snap-shots, obtained over a sufficient period of time, on the observation plane. The observations are carried out on a plane perpendicular to the inlet plane with a shared edge normal to the homogeneity direction. The method is applicable to all flows which display a direction of homogeneity such as cylinder wake flows, channel flow, mixing layer, and jet (axi-symmetric). The ability of the method is assessed with two synthetic data-sets, and three experimental PIV data-sets. A good reconstruction of the large-scale structures are observed for all cases. This study has been published in the journal "Experiments in Fluids" [21].

6.4. Visual servoing approach for fluid flow control

6.4.1. A state space representation for the closed-loop control of shear flows

Participants: Johan Carlier, Christophe Collewet.

The goal of this study is to develop a generic state representation for the closed-loop control of shear flows. We assume that the actuator acts at the boundaries. Our approach is based on a linearization of the Navier-Stokes equations around the desired state. Particular care was paid to the discrete approximation of the linear model to design a well-conditioned and accurate state matrix describing time evolution of disturbances evolving in parallel shear flow as long as these disturbances remain sufficiently small. A state matrix representation is obtained for the periodic channel flow and the spatially developing mixing layer flow. This approach has been validated through the representativity of our model in terms of linear stability. This work has been presented to the French Mechanics Congress CFM'2019 (https://hal.inria.fr/hal-02283161) [36].

6.4.2. Closed-loop control of a spatially developing shear layer

Participants: Christophe Collewet, Johan Carlier.

This study aims at controling 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, in term of magnitude as well as in term of spatial dispersion. 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 (magnitude and spatial dispersion) has been developed to lead to an adaptive control law. Simple conditions to ensure the local asymptotic stability of the whole scheme have been derived. This work has been also presented to the French Mechanics Congress CFM'2019 (https://hal.archives-ouvertes.fr/hal-02189111) [37].

6.4.3. Design of a DBD plasma actuator for closed-loop control

Participants: Johan Carlier, Christophe Collewet.

The goal of this study is to design a DBD plasma actuator for closed-loop control. This kind of actuator is widely used in the flow control community however, it is more appropriate to force a flow than to control it. Indeed, to control a flow under a closed-loop fashion, the action must be proportional to the control signal provided by the control law. It is unfortunately not the case with these actuators. We have modified the classical DBD plasma actuator so that the action is almost a linear fonction of the control signal. Our approach have been validated by a prototype and by first experiments.

6.5. Coupled models in hydrogeology

6.5.1. Reactive transport in multiphase flow

Participant: Jocelyne Erhel.

Groundwater resources are essential for life and society, and should be preserved from contamination. Pollutants are transported through the porous medium and a plume can propagate. Reactive transport models aims at simulating this dynamic contamination by coupling advection dispersion equations with chemistry equations. If chemistry is at thermodynamic equilibrium, then the system is a set of partial differential and algebraic equations (PDAE). Space discretization leads to a semi-discrete DAE system which should be discretized in time. An explicit time scheme allows an easy decoupling of transport and chemistry, but very small timesteps should be taken, leading to a very large CPU time. Therefore, an implicit time scheme is preferred, coupling transport and chemistry in a nonlinear system. The special structure of linearized systems can be used in preconditioned Newton-Krlyov methods in order to improve efficiency. Some experiments illustrate the methodology and show also the need for an adaptive timestep and a control of convergence in Newton's iterations.

This work was presented at a workshop [31].

6.5.2. Characterizations of Solutions in Geochemistry at equilibrium

Participant: Jocelyne Erhel.

Geochemistry at thermodynamic equilibrium involves aqueous reactions and mineral precipitation or dissolution. Quantities of solute species are assumed to be strictly positive, whereas those of minerals can vanish. The mathematical model is expressed as the minimization of Gibbs energy subject to positivity of mineral quantities and conservation of mass. Optimality conditions lead to a complementarity problem. We show that, in the case of a dilute solution, this problem can also be considered as optimality conditions of another minimization problem, subject to inequality constraints. This new problem is easier to handle, both from a theoretical and a practical point of view. Then we define a partition of the total quantities in the mass conservation equation. This partition builds a precipitation diagram such that a mineral is either precipitated or dissolved in each subset. We propose a symbolic algorithm to compute this diagram. Simple numerical examples illustrate our methodology.

This work was published in the journal Computational Geosciences [22] and presented at an international conference [38].

6.5.3. Mathematical models of kinetic reactions in geochemistry

Participants: Jocelyne Erhel, Bastien Hamlat.

In geochemistry, kinetic reactions can lead to the appearance or disappearance of minerals or gas. We defined two mathematical models based first on a differential inclusion system and second on a projected dynamical system. We proposed a regularization process for the first model and a projection algorithm for the second one.

This work, supported by IFPEN, was presented at a conference [39] and a workshop [32].

6.6. Sparse Linear solvers

6.6.1. Parallel GMRES

Participant: Jocelyne Erhel.

Sparse linear systems $Ax = b$ arise very often in computational science and engineering. Krylov methods are very efficient iterative methods, and restarted GMRES is a reference algorithm for non-symmetric systems. A first issue is to ensure a fast convergence, by preconditioning the system with a matrix M. Preconditioning must reduce the number of iterations, and be easy to solve. A second issue is to achieve high performance computing. The most time-consuming part in GMRES is to build an orthonormal basis V. With the Arnoldi process, many scalar products involve global communications. In order to avoid them, s-step methods have been designed to find a tradeoff between parallel performance and stability. Also, solving a system with the matrix M and for multiplying a vector by the matrix A should be efficient. A domain decomposition approach involves mainly local communications and is frequently used. A coarse grid correction, based on deflation for example, improves convergence. These techniques can be combined to provide fast convergence and fast parallel algorithms. Numerical results illustrate various issues and achievements.

This work was presented at an international conference (invited talk) [31].

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. 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.

7.1.2. 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. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. Comins'lab: SEACS : Stochastic modEl-dAta-Coupled representationS for the analysis, simulation and reconstruction of upper ocean dynamics

Participant: 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.

8.1.2. ANR BECOSE : Beyond Compressive Sensing: Sparse approximation algorithms for ill-conditioned inverse problems.

Participant: Dominique Heitz.

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.

8.1.3. IFPEN project

Participants: Jocelyne Erhel, Bastien Hamlat.

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. Contract with IFPEN (Institut FranÃ§ais du Pétrole et Energies Nouvelles). Duration: three years October 2016-September 2019. 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 use. This project focused on chemistry models, with kinetic reactions. We developed a mathematical tool, which can be embedded into a reactive transport code.

8.1.4. 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. We participated in a workshop on reactive transport (SITRAM), Pau, December 2019.

8.1.5. LEFE MANU: MSOM

Participants: Etienne Mémin, Long Li.

Title: Multiple Scale Ocean Model
Duration: From 2018 to 2021
Coordination: Bruno Deremble (CNRS LMD/ENS Paris)
Abstract: The objective of this project is to propose a numerical framework of a multiscale ocean model and to demonstrate its utility in the understanding of the interaction between the mean current and eddies.

8.2. International Initiatives

8.2.1. Inria International Partners

8.2.1.1. Informal International Partners

Imperial College, London (UK), Collaboration with Dan Crişan 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.

8.2.1.2. International Initiatives

MATH-GEO

Title: MATHematical methods for GEOphysical flows

International Partners (Institution - Laboratory - Researcher):

- Universidad de Buenos Aires (Argentina) - CIMA - Juan Ruiz
- Universidad de la Republica Uruguay (Uruguay) - IMFIA, INCO
- CMM (Chile) - Center for Mathematical Modeling - Axel Osses
- Universidad San Ignacio de Loyola (USIL) (Peru) - Faculty of Engineering Alejandro Paredes

Duration: 2018 - 2019

Start year: 2018 http://mathgeo.cima.fcen.uba.ar

Nonlinear processes, such as advection and turbulent mixing, play a central role in geophysical sciences. The theory of nonlinear dynamical systems provides a systematic way to study these phenomena. Its stochastic extension also forms the basis of modern data analysis techniques, predictability studies and data assimilation methods. Contributions in the field of Topology and Dynamics of Chaos include methods conceived to unveil the structure organizing flows in phase space, building the gap between data and l ow-dimensional modeling. Low-order models in climate dynamics are highly desirable, since they can provide solutions in cases where high-resolution numerical simulations cannot be implemented, as in short-term wind forecasting. At the same time, the procedure provides a tool-kit for model validation, emulation or inter-model comparison, with interesting prospects in all fields of oceanographic and atmospheric sciences, including climate detection and attribution. The strategy constitutes an unprecedented and promising perspective, offering an original approach to the subject, with mathematical concepts that are not necessarily widespread in the geophysics scientific community. This proposal gathers specialists with a know-how in the most challenging aspects of the focused research field: coherent structure detection in fluid flows for the exploration and interactive visualization of scientific data (LIMSI France), data assimilation and fluid motion analysis from image sequences (Inria Rennes), numerical models and data assimilation (CMM-Chile) stochastic models for climate dynamics with application to El Niño Ocean models (USIL-Peru), mathematical methods for weather and climate (CIMA-UBA & IMIT / IFAECI, Argentina), geophysical flows and dynamical systems (LMD France), mixing structures and Lagrangian analysis of multisatellite data (LOCEAN France), marine and estuarine hydrodynamic and water properties numerical models (INCO & IMFIA-Uruguay), in situ measurements of oceanographic conditions (CEBC France, in program with CNES France and CONAE Argentina), global modelling technique and topological characterization of flows (CORIA with CESBIO, France).

8.3. International Research Visitors

8.3.1. Visits of International Scientists

- 1 week visit of Alejandro Paredes Universidad San Ignacio de Loyola (USIL) (Peru) to work with Etienne Mémin
- 1 week visit of André Cavaleri (Instituto Tecnologico de Aeronautica, SP, Brésil) to work with Gilles Tissot
- 3 months visit of Ruediger Brecht (May to August), PhD student at Memorial University of Newfoundland, Canada, supported by funding from Inria-Mitacs Globalink.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Selection

Jocelyne Erhel is member of

- the international advisory committee of the parallel CFD conferences (Antalya, May 2019).
- the program committee of the workshop Visualization in Environmental Sciences 2019 (co-event of EuroVis).
- the scientific Committee of the workshop "Parallel solution methods for systems arising from PDEs" (Marseille, September 2019).
- the scientific Committee of the conference SimRace (IFPEN, Rueil-Malmaison, scheduled in December 2019 and postponed in 2020).

9.1.2. Journal

9.1.2.1. 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)

9.1.2.2. Reviewer - Reviewing Activities

Jocelyne Erhel: Reviewer for the journals Computational Geosciences, SISC, M2AN, JCAM, PARCO
Dominique Heitz: Reviewer for Exp. in Fluids, AMI Région Auvergne Rhone Alpes
Etienne Mémin: Reviewer for Tellus-A, Quat. J. of the Roy. Met. Soc., Journ. of Fluid Mech., Im. Vis. Comp., Exp. in Fluids, Journ. of Comp. Phys.
Gilles Tissot: Reviewer for Journal of Sound and Vibration, Fluid Dynamics Research.

9.1.3. Invited Talks

Werner Bauer

- European Numerical Mathematics and Advanced Applications Conference 2019, Egmond aan Zee, The Netherlands.
- Seminar talk at IRMAR, University of Rennes, France.
- Seminar talk at the Journée Rennes-Nantes d'Analyse 2019, University of Nantes, France.

Jocelyne Erhel

- Workshop on "Reactive transport modeling", IPGP, Paris, June 2019.
- Workshop on "Parallel Solution Methods for Systems Arising from PDEs", Marseille, France, September 2019.
- Workshop on "Advances in the SImulation of reactive flow and TRAnsport in porous Media (SITRAM)",

Etienne Mémin

- Workshop Big data, data assimilation, and uncertainty quantification, IHP, Paris (France), 12-15 November 2019
- Workshop on stochastic parameterizations and their use in data assimilation 1-5 July 2019 Imperial College London
- Equadif 2019, Leiden The Netherland, July 2019
- EGU, Vienna, Austria, April 2019
- Workshop on Conservation Principles, Data & Uncertainty in Atmosphere-Ocean Modelling", Potsdam, Germany April 2019

9.1.4. Leadership within the Scientific Community

- J. Erhel is scientific coordinator of the website Interstices (since June 2012). https://interstices.info.

9.1.5. Scientific Expertise

- J. Erhel is a member of the scientific council of IFPEN, since April 2016.

9.1.6. Research Administration

Jocelyne Erhel

- the Inria administrative commission (CAP) for researchers, 2016-2019.
- the maths thesis committee of IRMAR, 2017-2019.
- the selection committee for PhD grants of OSUR, 2019.

Dominique Heitz

- Responsible of the Irstea ACTA Team
- Member of Irstea OPAALE research unit Executive Committee
- Member of Pôle Cristal scientific council
- D. Heitz is a member of scientific council of CSTB's Jules Verne Wind Tunnel

Roger Lewandowski

- President du Comité de liaison du groupe GAMNI-SMAI
- President of the Blaise Pascal award jury
- President of GAMNI-SMAI PhD thesis award
- Corresponding person of the SMAI in Rennes
- Responsible of the group "Mathematical modeling" of IRMAR
- Member of the scientific council of IRMAR,
- Member of Mathematical teaching council of U. Rennes I
- Member of the scientific council of the Henri Lebesgue Centre

Etienne Mémin

- Member of the scientific council of LEFE-MANU action of CNRS INSU
- Member of the comity GAMNI-SMAI

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Licence: Jocelyne Erhel, Optimisation, 12h, 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: Roger Lewandowski, Euler and the Navier-Stokes equations, M2, master Â« fondamental mathematics Â».

Master : Etienne Mémin, Analyse du mouvement, Master Informatique, 15h, niveau M2, Université de Rennes 1.

Master : Etienne Mémin, Vision par ordinateur , 15h, niveau M2, ESIR Université de Rennes 1.

Master : Etienne Mémin, Motion analysis , 9h, Master 2 SISEA Université de Rennes 1.

Master : Gilles Tissot, mathematics for acoustics, 20h, niveau M1, Université du Mans.

9.2.2. Supervision

PhD in progress: Bastien Hamlat, University of Rennes 1, October 2016, co-advisors Jocelyne Erhel and A. Michel.

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, Gilles Tissot.

PhD in progress : Dinh Duong Nguyen, Regular and singular solutions of Navier-Stokes equations with eddy viscosity, started September 2017 , Roger Lewandowski.

PhD in progress : Robin Billard, Modelling of non-conventional perforated acoustic liners, Université du Mans, started December 2017, Gilles Tissot.

9.2.3. Juries

Jocelyne Erhel

- Etienne Ahusborde, HdR, Univ. Strasbourg (rapporteur)
- Benoit Pinier, PhD, Univ. Rennes (examinatrice)
- Quentin Tournois, PhD, Univ. Rennes (examinatrice)

Etienne Mémin

- Anthony Fillion, Ecoles des Ponts, U. Paris-Est (Rapporteur)
- Alban Farchi, Ecoles des Ponts, U. Paris-Est (Examinateur)
- Arthur Pajot, Sorbonne Université, (Rapporteur)

9.3. Popularization

9.3.1. Education

Jocelyne Erhel

- présidente du jury du rallye de mathématiques du CNED, since 2017.

9.3.2. Interventions

Jocelyne Erhel

- conference about spreading epidemics, "les amphis lycéens 2018-2019", March 2019.
- conference about spreading epidemics, "les amphis lycéens 2019-2020", December 2019. [46].

9.3.3. Creation of media or tools for science outreach

Jocelyne Erhel

- was scientific coordinator of the website Interstices (June 2012 - September 2019). She is now member of the editorial board, from October 2019.

10. 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^{o} 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^{o} 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^{o} 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^{o} 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^{o} 1, p. 81–103, http://www.irisa.fr/fluminance/publi/papers/Papadakis-Memin-JMIV07.pdf

[12] V. RESSEGUIER, E. MÉMIN, B. CHAPRON.*Geophysical flows under location uncertainty, Part I Random transport and general models*, in "Geophys. & Astro. Fluid Dyn.", 2017, vol. 111, n^{o} 3, p. 149-176

[13] V. RESSEGUIER, E. MÉMIN, B. CHAPRON.*Geophysical flows under location uncertainty, Part II Quasi-geostrophy and efficient ensemble spreading*, in "Geophysical and Astrophysical Fluid Dynamics", April 2017, vol. 111, n^{o} 3, p. 177-208 [*DOI :* 10.1080/03091929.2017.1312101], https://hal.inria.fr/hal-01391476

[14] 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^{o} 3, p. 209-227 [*DOI :* 10.1080/03091929.2017.1312102], https://hal.inria.fr/hal-01391484

[15] V. RESSEGUIER, E. MÉMIN, D. HEITZ, B. CHAPRON.*Stochastic modelling and diffusion modes for proper orthogonal decomposition models and small-scale flow analysis*, in "J. Fluid Mech.", 2017, vol. 828, 29

[16] 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^{o} 5, p. 1909-1926

Publications of the year

Doctoral Dissertations and Habilitation Theses

[17] B. PINIER.*Application of the similartity theory of turbulence to the ocean-atmosphere interface*, Université Rennes 1, February 2019, https://tel.archives-ouvertes.fr/tel-02057601

Articles in International Peer-Reviewed Journal

[18] L. C. BERSELLI, T. ILIESCU, B. KOC, R. LEWANDOWSKI.*Long-Time Reynolds Averaging of Reduced Order Models for Fluid Flows: Preliminary Results*, in "Mathematics in Engineering", 2020, vol. 2, n^{o} 1, p. 1-25, https://arxiv.org/abs/1901.04903 - 20 pages, 1 figure, 9 tables [*DOI :* 10.3934/MINE.2020001], https://hal.archives-ouvertes.fr/hal-01979939

[19] L. C. BERSELLI, R. LEWANDOWSKI.*On the Reynolds time-averaged equations and the long-time behavior of Leray-Hopf weak solutions, with applications to ensemble averages*, in "Nonlinearity", October 2019, vol.

32, n° 11, p. 4579-4608, https://arxiv.org/abs/1801.08721 - 33 pages [*DOI :* 10.1088/1361-6544/AB32BC], https://hal.archives-ouvertes.fr/hal-01695374

[20] R. Brecht, W. Bauer, A. Bihlo, F. Gay-Balmaz, S. MacLachlan.*Variational integrator for the rotating shallow-water equations on the sphere*, in "Quarterly Journal of the Royal Meteorological Society", March 2019, vol. 145, n° 720, p. 1070-1088, https://arxiv.org/abs/1808.10507 [*DOI :* 10.1002/QJ.3477], https://hal.archives-ouvertes.fr/hal-02414164

[21] P. Chandramouli, E. Mémin, D. Heitz, L. Fiabane.*Fast 3D flow reconstructions from 2D cross-plane observations*, in "Experiments in Fluids", February 2019, p. 1-27 [*DOI :* 10.1007/S00348-018-2674-1], https://hal.archives-ouvertes.fr/hal-02012453

[22] J. Erhel, T. Migot.*Characterizations of Solutions in Geochemistry: Existence, Uniqueness and Precipitation Diagram*, in "Computational Geosciences", 2019, vol. 23, n° 3, p. 523-535 [*DOI :* 10.1007/S10596-018-9800-2], https://hal.inria.fr/hal-01584490

[23] M. Khalid, L. Pénard, E. Mémin.*Optical Flow For Image-Based River Velocity Estimation*, in "Flow Measurement and Instrumentation", 2019, vol. 65, p. 110-121 [*DOI :* 10.1016/J.FLOWMEASINST.2018.11.009], https://hal.inria.fr/hal-01930498

[24] R. Lewandowski.*Navier-Stokes equations in the whole space with an eddy viscosity*, in "Journal of Mathematical Analysis and Applications", 2019, vol. 478, n° 2, p. 698-742, https://arxiv.org/abs/1705.11043 [*DOI :* 10.1016/J.JMAA.2019.05.051], https://hal.archives-ouvertes.fr/hal-01531260

[25] B. Pinier, E. Mémin, S. Laizet, R. Lewandowski.*A stochastic flow approach to model the mean velocity profile of wall-bounded flows*, in "Physical Review E ", 2019, vol. 99, n° 6, 063101, 27 pages [*DOI :* 10.1103/PHYSREVE.99.063101], https://hal.inria.fr/hal-01947662

[26] R. Schuster, D. Heitz, P. Georgeault, E. Mémin.*On-site airflow measurement of a laboratory fume hood using customized large-scale image-based velocimetry*, in "Indoor and Built Environment", July 2019, p. 1-10 [*DOI :* 10.1177/1420326X19865928], https://hal.archives-ouvertes.fr/hal-02419135

[27] G. Tissot, R. Billard, G. Gabard.*Optimal cavity shape design for acoustic liners using Helmholtz equation with visco-thermal losses*, in "Journal of Computational Physics", 2020, vol. 402, 109048 [*DOI :* 10.1016/J.JCP.2019.109048], https://hal.inria.fr/hal-02335784

[28] G. Wimmer, C. J. Cotter, W. Bauer.*Energy conserving upwinded compatible finite element schemes for the rotating shallow water equations*, in "Journal of Computational Physics", January 2020, vol. 401, 109016, https://arxiv.org/abs/1901.06349 [*DOI :* 10.1016/J.JCP.2019.109016], https://hal.archives-ouvertes.fr/hal-02419835

[29] Y. Yang, E. Mémin.*Estimation of physical parameters under location uncertainty using an Ensemble^2-Expectation-Maximization algorithm*, in "Quarterly Journal of the Royal Meteorological Society", January 2019, vol. 145, n° 719, p. 418-433 [*DOI :* 10.1002/QJ.3438], https://hal.inria.fr/hal-01944730

Articles in National Peer-Reviewed Journal

[30] R. Schuster, D. Heitz, E. Mémin.*Visualisation et mesure du flux d'aspiration d'une sorbonne*, in "Salles Propres", May 2019, n° 119, p. 46-50, https://hal.archives-ouvertes.fr/hal-02418674

Invited Conferences

[31] J. ERHEL.*Numerical tools for simulating large scale reactive transport systems*, in "2019 - Parallel Solution Methods for Systems Arising from PDEs", Marseille, France, September 2019, https://hal.inria.fr/hal-02404231

[32] A. MICHEL, B. HAMLAT, J. ERHEL, T. FANEY.*Multiphase chemical kinetics modeling for aqueous-minerals reactive systems*, in "SITRAM 2019 - Advances in the SImulation of reactive flow and TRAnsport in porous Media", Pau, France, December 2019, 1, https://hal.inria.fr/hal-02404259

International Conferences with Proceedings

[33] R. BILLARD, G. TISSOT, G. GABARD, M. VERSAEVEL, J.-P. GROBY.*A mode-matching model for the impedance of perforated plate liners*, in "Aeroacoustics 2019 - 25th AIAA/CEAS Aeroacoustics Conference", Delft, Netherlands, American Institute of Aeronautics and Astronautics, May 2019, p. 1-12 [*DOI :* 10.2514/6.2019-2702], https://hal.inria.fr/hal-02335914

[34] V. RESSEGUIER, M. LADVIG, A. M. PICARD, E. MÉMIN, R. BOUAIDA, B. CHAPRON.*Quantifying the uncertainties introduced by dimension reduction in fluid dynamics*, in "UNCECOMP 2019 - 3rd International Conference on Uncertainty Quantification in Computational Sciences and Engineering", Hersonissos, Greece, June 2019, p. 1-21, https://hal.inria.fr/hal-02165809

[35] G. TISSOT, G. GABARD.*Optimal cavity shape design for acoustic liners using Helmholtz equation with visco-thermal losses*, in "Aeroacoustics 2019 - 25th AIAA/CEAS Aeroacoustics Conference", Delft, Netherlands, American Institute of Aeronautics and Astronautics, May 2019, p. 1-11 [*DOI :* 10.2514/6.2019-2471], https://hal.inria.fr/hal-02335834

Conferences without Proceedings

[36] J. CARLIER, C. COLLEWET.*Une représentation d'état pour la commande en boucle fermée des écoulements cisaillés*, in "CFM 2019 - 24e Congrès Français de Mécanique", Brest, France, August 2019, p. 1-17, https://hal.inria.fr/hal-02283161

[37] C. COLLEWET, J. CARLIER.*Commande en boucle fermée d'une couche de mélange soumise à des perturbations exogènes*, in "CFM 2019 - 24e Congrès Français de Mécanique", Brest, France, August 2019, p. 1-15, https://hal.archives-ouvertes.fr/hal-02189111

[38] J. ERHEL, T. MIGOT.*Optimization problems in geochemistry*, in "ETNA 2019 - 25th Recent Advances in Scientific Computation", Santa Margherita di Pula, Italy, May 2019, https://hal.inria.fr/hal-02404247

[39] B. HAMLAT, T. MIGOT, J. ERHEL, A. MICHEL.*A Differential Inclusion Approach to Mineral Precipitation-Dissolution Reactions in Geochemistry*, in "WCGO 2019 - 6th World Congress on Global Optimization", Rennes, France, July 2019, https://hal.inria.fr/hal-02404438

[40] S. HERPIN, D. HEITZ, P. LOISEL, P. GEORGEAULT.*Large-scale Stereo-PIV measurement of the flow inside an urban street canyon in outdoor conditions*, in "ISPIV2019 - 13th International Symposium on Particle Image Velocimetry", Munich, Germany, July 2019, p. 1-9, https://hal.archives-ouvertes.fr/hal-02330267

[41] V. RESSEGUIER, M. LADVIG, A. M. PICARD, E. MÉMIN, R. BOUAIDA, B. CHAPRON.*Stochastic reduced order model for real-time unsteady flow estimation*, in "WESC 2019 - Web Wind Energy Science Conference", Cork, Ireland, June 2019, p. 1-23, https://hal.archives-ouvertes.fr/hal-02160160

[42] V. RESSEGUIER, L. LI, G. JOUAN, P. DERIAN, E. MÉMIN, B. CHAPRON.*Dynamics under location uncertainty and other energy-related stochastic subgrid schemes*, in "2019 - Workshop Conservation Principles, Data and Uncertainty in Atmosphere-Ocean Modelling", Potsdam, Germany, April 2019, p. 1-70, https://hal.archives-ouvertes.fr/hal-02103233

[43] *Best Paper*
R. SCHUSTER, D. HEITZ, E. MÉMIN.*Visualisation et mesure du flux d'aspiration d'une Sorbonne*, in "ContaminExpert 2019", Paris, France, March 2019, p. 1-8, https://hal.archives-ouvertes.fr/hal-02330348.

[44] Y. YANG, D. HEITZ, E. MÉMIN.*Lagrangian Particle Image Velocimetry*, in "ISPIV2019 - 13th International Symposium on Particle Image Velocimetry", Munich, Germany, July 2019, p. 1-9, https://hal.archives-ouvertes.fr/hal-02330256

Scientific Books (or Scientific Book chapters)

[45] W. BAUER, F. GAY-BALMAZ.*Variational Discretization Framework for Geophysical Flow Models*, in "GSI 2019: Geometric Science of Information", Springer, August 2019, p. 523-531 [*DOI :* 10.1007/978-3-030-26980-7_54], https://hal.archives-ouvertes.fr/hal-02419889

Scientific Popularization

[46] J. ERHEL.*Les pics d'épidémie mis en équations*, in "Les amphis lycéens de l'université de Rennes 1", Rennes, France, 2019, https://hal.inria.fr/hal-02404288

Other Publications

[47] C. AMROUCHE, L. C. BERSELLI, R. LEWANDOWSKI, D. DUONG NGUYEN.*Turbulent flows as generalized Kelvin-Voigt materials: modeling and analysis*, August 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02267193

[48] W. BAUER, J. BEHRENS, C. J. COTTER.*A structure-preserving approximation of the discrete split rotating shallow water equations*, December 2019, https://arxiv.org/abs/1912.10335 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02423620

[49] W. BAUER, J. BEHRENS, C. J. COTTER.*A structure-preserving split finite element discretization of the rotating shallow water equations in split Hamiltonian form*, February 2019, working paper or preprint, https://hal.inria.fr/hal-02020379

[50] W. BAUER, P. CHANDRAMOULI, B. CHAPRON, L. LI, E. MÉMIN.*Deciphering the role of small-scale inhomogeneity on geophysical flow structuration: a stochastic approach*, December 2019, working paper or preprint, https://hal.inria.fr/hal-02398521

[51] W. BAUER, F. GAY-BALMAZ.*Variational integrators for soundproof models on arbitrary triangular C-grids*, January 2019, working paper or preprint, https://hal.inria.fr/hal-01970335

[52] C. CINTOLESI, E. MÉMIN.*Pseudo-stochastic simulation of turbulent channel flows with near-wall modelling*, February 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02044818

[53] C. CINTOLESI, E. MÉMIN.*Stochastic modelling of turbulent flows for numerical simulations*, February 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02044809

[54] J. ERHEL, M. OUMOUNI, G. PICHOT, F. SCHOEFS.*Analysis of continuous spectral method for sampling stationary Gaussian random fields*, April 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02109037

[55] R. LEWANDOWSKI.*On a one dimensional turbulent boundary layer model*, November 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02134034

References in notes

[56] R. ADRIAN.*Particle imaging techniques for experimental fluid mechanics*, in "Annal Rev. Fluid Mech.", 1991, vol. 23, p. 261-304

[57] T. BEWLEY.*Flow control: new challenges for a new Renaissance*, in "Progress in Aerospace Sciences", 2001, vol. 37, p. 21–58

[58] S. BEYOU, A. CUZOL, S. GORTHI, E. MÉMIN.*Weighted Ensemble Transform Kalman Filter for Image Assimilation*, in "TellusA", January 2013, vol. 65, n^o 18803

[59] S. CAI, E. MÉMIN, P. DÉRIAN, C. XU.*Motion Estimation under Location Uncertainty for Turbulent Fluid Flow*, in "Exp. in Fluids", 2017, vol. 59, n^o 8

[60] 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

[61] T. CORPETTI, E. MÉMIN.*Stochastic Uncertainty Models for the Luminance Consistency Assumption*, in "IEEE Trans. Image Processing", 2012, vol. 21, n^o 2, p. 481–493

[62] 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^o 10, p. 143–162

[63] J. FAVIER.*Contrôle d'écoulements : approche expérimentale et modélisation de dimension réduite*, Institut National Polytechnique de Toulouse, 2007

[64] 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

[65] 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^o 2

[66] 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

[67] B. HORN, B. SCHUNCK.*Determining Optical Flow*, in "Artificial Intelligence", August 1981, vol. 17, n^{o} 1-3, p. 185–203

[68] F.-X. LE DIMET, O. TALAGRAND.*Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects*, in "Tellus", 1986, n^{o} 38A, p. 97–110

[69] J. LIONS.*Optimal Control of Systems Governed by Partial Differential Equations*, Springer-Verlag, 1971

[70] 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

[71] 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^{o} 2, p. 810–826

[72] 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 GALLINETTE

Gallinette: developing a new generation of proof assistants

IN COLLABORATION WITH: Laboratoire des Sciences du numerique de Nantes

IN PARTNERSHIP WITH:
IMT Atlantique Bretagne-Pays de la Loire

Université Nantes

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Proofs and Verification

Table of contents

Project-Team GALLINETTE

Creation of the Team: 2017 May 01, updated into Project-Team: 2018 June 01

Keywords:

Computer Science and Digital Science:

A2.1.1. - Semantics of programming languages
A2.1.2. - Imperative programming
A2.1.3. - Object-oriented programming
A2.1.4. - Functional programming
A2.1.11. - Proof languages
A2.2.3. - Memory management
A2.4.3. - Proofs
A7.2.3. - Interactive Theorem Proving
A7.2.4. - Mechanized Formalization of Mathematics
A8.4. - Computer Algebra

Other Research Topics and Application Domains:

B6.1. - Software industry

1. Team, Visitors, External Collaborators

Research Scientists

Nicolas Tabareau [Team leader, Inria, Researcher, HDR]
Assia Mahboubi [Inria, Researcher]
Guillaume Munch-Maccagnoni [Inria, Researcher]
Pierre-Marie Pédrot [Inria, Researcher]

Faculty Members

Julien Cohen [Univ de Nantes, Associate Professor]
Rémi Douence [IMT Atlantique, Associate Professor, HDR]
Hervé Grall [IMT Atlantique, Associate Professor]
Guilhem Jaber [Univ de Nantes, Associate Professor]

Technical Staff

Simon Boulier [Inria, Engineer]

PhD Students

Antoine Allioux [Inria]
Meven Bertrand [Univ de Nantes, from Sep 2019]
Gaetan Gilbert [IMT Atlantique]
Ambroise Lafont [IMT Atlantique]
Xavier Montillet [Univ de Nantes]
Loic Pujet [Univ de Nantes, from Sep 2019]
Theo Winterhalter [Univ de Nantes]
Igor Zhirkov [Armines, until Sep 2019]

Post-Doctoral Fellows

Eric Finster [Inria, until Jan 2019]
Marie Kerjean [Inria]
Maxime Lucas [Inria]

Kenji Maillard [Inria, from Dec 2019]
Étienne Miquey [Inria, until Nov 2019]
Pierre Vial [Inria]

2. Overall Objectives

2.1. Overall Objectives

The EPI Gallinette aims at developing a new generation of proof assistants, with the belief that practical experiments must go in pair with foundational investigations:

- The goal is to advance proof assistants both as certified programming languages and mechanised logical systems. Advanced programming and mathematical paradigms must be integrated, notably dependent types and effects. The distinctive approach is to implement new programming and logical paradigms on top of Coq by considering the latter as a target language for compilation.
- The aim of foundational investigations is to extend the boundaries of the Curry-Howard correspondence. It is seen both as providing foundations for programming languages and logic, and as a purveyor of techniques essential to the development of proof assistants. Under this perspective, the development of proof assistants is seen as a total experiment using the correspondence in every aspect: programming languages, type theory, proof theory, rewriting and algebra.

3. Research Program

3.1. Scientific Context

Software quality is a requirement that is becoming more and more prevalent, by now far exceeding the traditional scope of embedded systems. The development of tools to construct software that respects a given specification is a major challenge facing computer science. *Proof assistants* such as Coq [49] provide a formal method whose central innovation is to produce *certified programs* by transforming the very activity of programming. Programming and proving are merged into a single development activity, informed by an elegant but rigid mathematical theory inspired by the correspondence between programming, logic and algebra: the *Curry-Howard correspondence*. For the certification of programs, this approach has shown its efficiency in the development of important pieces of certified software such as the C compiler of the CompCert project [78]. The extracted CompCert compiler is reliable and efficient, running only 15% slower than GCC 4 at optimisation level 2 (`gcc -O2`), a level of optimisation that was considered before to be highly unreliable.

Proof assistants can also be used to *formalise mathematical theories*: they not only provide a means of representing mathematical theories in a form amenable to computer processing, but their internal logic provides a language for reasoning about such theories. In the last decade, proof assistants have been used to verify extremely large and complicated proofs of recent mathematical results, sometimes requiring either intensive computations [60], [64] or intricate combinations of a multitude of mathematical theories [59]. But formalised mathematics is more than just proof checking and proof assistants can help with the organisation mathematical knowledge or even with the discovery of new constructions and proofs.

Unfortunately, the rigidity of the theory behind proof assistants impedes their expressiveness both as programming languages and as logical systems. For instance, a program extracted from Coq only uses a purely functional subset of OCaml, leaving behind important means of expression such as side-effects and objects. Limitations also appears in the formalisation of advanced mathematics: proof assistants do not cope well with classical axioms such as excluded middle and choice which are sometimes used crucially. The fact of the matter is that the development of proof assistants cannot be dissociated from a reflection on the nature of programs and proofs coming from the Curry-Howard correspondence. In the EPC Gallinette, we propose to address several drawbacks of proof assistants by pushing the boundaries of this correspondence.

In the 1970's, the Curry-Howard correspondence was seen as a perfect match between functional programs, intuitionistic logic, and Cartesian closed categories. It received several generalisations over the decades, and now it is more widely understood as a fertile correspondence between computation, logic, and algebra. Nowadays, the view of the Curry-Howard correspondence has evolved from a perfect match to a collection of theories meant to explain similar structures at work in logic and computation, underpinned by mathematical abstractions. By relaxing the requirement of a perfect match between programs and proofs, and instead emphasising the common foundations of both, the insights of the Curry-Howard correspondence may be extended to domains for which the requirements of programming and mathematics may in fact be quite different.

Consider the following two major theories of the past decades, which were until recently thought to be irreconcilable:

- **(Martin-Löf) Type theory:** introduced by Martin-Löf in 1971, this formalism [85] is both a programming language and a logical system. The central ingredient is the use of *dependent types* to allow fine-grained invariants to be expressed in program types. In 1985, Coquand and Huet developed a similar system called the *calculus of constructions*, which served as logical foundation of the first implementation of Coq. This kind of systems is still under active development, especially with the recent advent of homotopy type theory (HoTT) [107] which gives a new point of view on types and the notion of equality in type theory.
- **The theory of effects:** starting in the 1980's, Moggi [90] and Girard [57] put forward monads and co-monads as describing various compositional notions of computation. In this theory, programs can have side-effects (state, exceptions, input-output), logics can be non-intuitionistic (linear, classical), and different computational universes can interact (modal logics). Recently, the safe and automatic management of resources has also seen a coming of age (Rust, Modern C++) confirming the importance of linear logic for various programming concepts. It is now understood that the characteristic feature of the theory of effects is sensitivity to *evaluation order*, in contrast with type theory which is built around the assumption that evaluation order is irrelevant.

We now outline a series of scientific challenges aimed at understanding of type theory, effects, and their combination.

More precisely, three key axes of improvement have been identified:

1. Making the notion of equality closer to what is usually assumed when doing proofs on black board, with a balance between irrelevant equality for simple structures and equality up-to equivalences for more complex ones (Section 3.2). Such a notion of equality should allow one to implement traditional model transformations that enhance the logical power of the proof assistant using distinct compilation phases.
2. Advancing the foundations of effects within the Curry-Howard approach. The objective is to pave the way for the integration of effects in proof assistants and to prototype the corresponding implementation. This integration should allow for not only certified programming with effects, but also the expression of more powerful logics (Section 3.3).
3. Making more programming features (notably, object polymorphism) available in proof assistants, in order to scale to practical-sized developments. The objective is to enable programming styles closer to common practices. One of the key challenges here is to leverage gradual typing to dependent programming (Section 3.4).

To validate the new paradigms, we propose in Section 3.5 three particular application fields in which members of the team already have a strong expertise: code refactoring, constraint programming and symbolic computation.

3.2. Enhance the computational and logical power of proof assistants

The democratisation of proof assistants based on type theory has likely been impeded one central problem: the mismatch between the conception of equality in mathematics and its formalisation in type theory. Indeed, some basic principles that are used implicitly in mathematics—such as Church's principle of propositional extensionality, which says that two propositions are equal when they are logically equivalent—are not derivable in type theory. Even more problematically, from a computer science point of view, the basic concept of two functions being equal when they are equal at every "point" of their domain is also not derivable: rather, it must be added as an additional axiom. Of course, these principles are consistent with type theory so that working under the corresponding additional assumptions is safe. But the use of these assumptions in a definition potentially clutters its computational behaviour: since axioms are computational black boxes, computation gets stuck at the points of the code where they have been used.

We propose to investigate how expressive logical transformations such as forcing [70] and sheaf construction might be used to enhance the computational and logical power of proof assistants—with a particular emphasis on their implementation in the Coq proof assistant by the means of effective translations (or compilation phases). One of the main topics of this task, in connection to the ERC project CoqHoTT, is the integration in Coq of new concepts inspired by homotopy type theory [107] such as the univalence principle, and higher inductive types.

3.2.1. A definitional proof-irrelevant version of Coq.

In the Coq proof assistant, the sort $\mathbf{Prop}$ stands for the universe of types which are propositions. That is, when a term P has type $\mathbf{Prop}$, the only relevant fact is whether P is inhabited (that is true) or not (that is false). This property, known as *proof irrelevance,* can be expressed formally as: $\forall x\ y : P, x = y$. Originally, the *raison d'être* of the sort $\mathbf{Prop}$ was to characterise types with no computational meaning with the intention that terms of such types could be erased upon extraction. However, the assumption that every element of $\mathbf{Prop}$ should be proof irrelevant has never been integrated to the system. Indeed, in Coq, proof irrelevance for the sort $\mathbf{Prop}$ is not incorporated into the theory: it is only compatible with it, in the sense that its assumption does not give rise to an inconsistent theory. In fact, the exact status of the sort $\mathbf{Prop}$ in Coq has never been entirely clarified, which explains in part this lack of integration. Homotopy type theory brings fresh thinking on this issue and suggests turning $\mathbf{Prop}$ into the collection of terms that a certain static inference procedure tags as proof irrelevant. The goal of this task is to integrate this insight in the Coq system and to implement a definitional proof-irrelevant version of the sort $\mathbf{Prop}$.

3.2.2. Extend the Coq proof assistant with a computational version of univalence

The univalence principle is becoming widely accepted as a very promising avenue to provide new foundations for mathematics and type theory. However, this principle has not yet been incorporated into a proof assistant. Indeed, the very mathematical structures (known as ∞-groupoids) motivating the theory remain to this day an active area of research. Moreover, a correct and decidable type checking procedure for the whole theory raises both computational complexity and logical coherence issues. Observational type theory [32], as implemented in Epigram, provides a first-stage approximation to homotopy type theory, but only deals with functional extensionality and does not capture univalence. Coquand and his collaborators have obtained significant results on the computational meaning of univalence using cubical sets [39], [45]. Bickford has initiated a promising formalisation work [0] in the NuPRL system. However, a complete formalisation in intensional type theory remains an open problem.

Hence a major objective is to achieve a complete internalisation of univalence in intensional type theory, including an integration to a new version of Coq. We will strive to keep compatibility with previous versions, in particular from a performance point of view. Indeed, the additional complexity of homotopy type theory should not induce an overhead in the type checking procedure used by the software if we want our new framework to become rapidly adopted by the community. Concretely, we will make sure that the compilation time of Coq's Standard Library will be of the same order of magnitude.

3.2.3. Extend the logical power of type theory without axioms in a modular way

Extending the power of a logic using model transformations (*e.g.,* forcing transformation [71], [70] or the sheaf construction [100]) is a classic topic of mathematical logic [46], [76]. However, these ideas have not

[0] http://www.nuprl.org/wip/Mathematics/cubical!type!theory/index.html

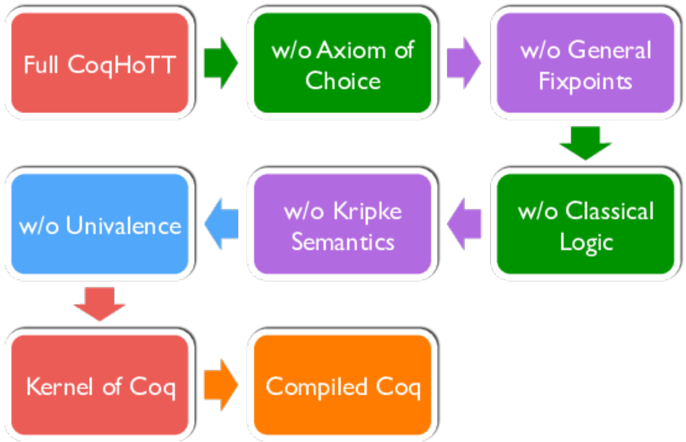


Figure 1. Multiple compilation phases to increase the logical and computational power of Coq.

been much investigated in the setting of type theory, even though they may provide a useful framework for extending the logical power of proof assistant in a modular way. There is a good reason for this: with a syntactic notion of equality, the underlying structure of type theory does not conform to the structure of topos used in mathematical logic. A direct incorporation of the standard techniques is therefore not possible. However, a univalent notion of equality brings type theory closer to the required algebraic structure, as it corresponds to the notion of ∞-topos recently studied by Lurie [83]. The goal of this task is to revisit model transformations in the light of the univalence principle, and to obtain in this way new internal transformations in type theory which can in turn be seen as compilation phases. The general notion of an internal syntactical translation has already been investigated in the team [40].

3.2.4. Methodology: Extending type theory with different compilation phases

The Gallinette project advocates the use of distinct compilation phases as a methodology for the design of a new generation of proof assistants featuring modular extensions of a core logic. The essence of a compiler is the separation of the complexity of a translation process into modular stages, and the organization of their re-composition. This idea finds a natural application in the design of complex proof assistants (Figure 1). For instance, the definition of type classes in Coq follows this pattern, and is morally given by the means of a translation into a type-class free kernel. More recently, a similar approach by compilation stages, using the forcing transformation, was used to relax the strict positivity condition guarding inductive types [71], [70]. We believe that this flavour of compilation-based strategies offers a promising direction of investigation for the propose of defining a decidable type checking algorithm for HoTT.

3.3. Semantic and logical foundations for effects in proof assistants based on type theory

We propose the incorporation of effects in the theory of proof assistants at a foundational level. Not only would this allow for certified programming with effects, but it would moreover have implications for both semantics and logic.

We mean *effects* in a broad sense that encompasses both Moggi's monads [90] and Girard's linear logic [57]. These two seminal works have given rise to respective theories of effects (monads) and resources (co-monads). Recent advances, however have unified these two lines of thought: it is now clear that the defining feature of effects, in the broad sense, is sensitivity to evaluation order [79], [50].

In contrast, the type theory that forms the foundations of proof assistants is based on pure λ calculus and is built on the assumption that evaluation order is irrelevant. Evaluation order is therefore the blind spot of type theory. In Moggi [91], integrating the dependent types of type theory with monads is *"the next difficult step [...] currently under investigation"*.

Any realistic program contains effects: state, exceptions, input-output. More generally, evaluation order may simply be important for complexity reasons. With this in mind, many works have focused on certified programming with effects: notably Ynot [95], and more recently F* [105] and Idris [41], which propose various ways for encapsulating effects and restricting the dependency of types on effectful terms. Effects are either specialised, such as the monads with Hoare-style pre- and post-conditions found in Ynot or F*, or more general, such as the algebraic effects implemented in Idris. But whereas there are several experiments and projects pursuing the certification of programs with effects, each making its own choices on how effects and dependency should be merged, there is on the other hand a deficit of logical and semantic investigations.

We propose to develop the foundations of a type theory with effects taking into account the logical and semantic aspects, and to study their practical and theoretical consequences. A type theory that integrates effects would have logical, algebraic and computational implications when viewed through the Curry-Howard correspondence. For instance, effects such as control operators establish a link with classical proof theory [62]. Indeed, control operators provide computational interpretations of type isomorphisms such as $A \cong \neg\neg A$ and $\neg\forall x A \cong \exists x \neg A$ (e.g. [92]), whereas the conventional wisdom of type theory holds that such axioms are non-constructive (this is for instance the point of view that has been advocated so far in homotopy type theory [107]). Another example of an effect with logical content is state (more precisely memoization) which is used to provide constructive content to the classical dependent axiom of choice [38], [74], [66]. In the long term, a whole body of literature on the constructive content of classical proofs is to be explored and integrated, providing rich sources of inspiration: Kohlenbach's proof mining [73] and Simpson's reverse mathematics [103], for instance, are certainly interesting to investigate from the Curry-Howard perspective.

The goal is to develop a type theory with effects that accounts both for practical experiments in certified programming, and for clues from denotational semantics and logical phenomena, in a unified setting.

3.3.1. Models for integrating effects with dependent types

A crucial step is the integration of dependent types with effects, a topic which has remained *"currently under investigation"* [91] ever since the beginning. The difficulty resides in expressing the dependency of types on terms that can perform side-effects during the computation. On the side of denotational semantics, several extensions of categorical models for effects with dependent types have been proposed [29], [108] using axioms that should correspond to restrictions in terms of expressivity but whose practical implications, however, are not immediately transparent. On the side of logical approaches [66], [67], [77], [89], one first considers a drastic restriction to terms that do not compute, which is then relaxed by semantic means. On the side of systems for certified programming such as F*, the type system ensures that types only depend on pure and terminating terms.

Thus, the recurring idea is to introduce restrictions on the dependency in order to establish an encapsulation of effects. In our approach, we seek a principled description of this idea by developing the concept of *semantic value* (thunkables, linears) which arose from foundational considerations [56], [102], [93] and whose relevance was highlighted in recent works [80], [99]. The novel aspect of our approach is to seek a proper extension of type theory which would provide foundations for a classical type theory with axiom of choice in the style of Herbelin [66], but which moreover could be generalised to effects other than just control by exploiting an abstract and adaptable notion of semantic value.

3.3.2. Intuitionistic depolarisation

In our view, the common idea that evaluation order does not matter for pure and termination computations should serve as a bridge between our proposals for dependent types in the presence of effects and traditional type theory. Building on the previous goal, we aim to study the relationship between semantic values, purity, and parametricity theorems [101], [58]. Our goal is to characterise parametricity as a form of intuitionistic *depolarisation* following the method by which the first game model of full linear logic was given (Melliès [86], [87]). We have two expected outcomes in mind: enriching type theory with intensional content without losing its properties, and giving an explanation of the dependent types in the style of Idris and F* where purity- and termination-checking play a role.

3.3.3. Developing the rewriting theory of calculi with effects

An integrated type theory with effects requires an understanding of evaluation order from the point of view of rewriting. For instance, rewriting properties can entail the decidability of some conversions, allowing the automation of equational reasoning in types [27]. They can also provide proofs of computational consistency (that terms are not all equivalent) by showing that extending calculi with new constructs is conservative [104]. In our approach, the λ-calculus is replaced by a calculus modelling the evaluation in an abstract machine [51]. We have shown how this approach generalises the previous semantic and proof-theoretic approaches [33], [79], [81], and overcomes their shortcomings [94].

One goal is to prove computational consistency or decidability of conversions purely using advanced rewriting techniques following a technique introduced in [104]. Another goal is the characterisation of weak reductions: extensions of the operational semantics to terms with free variables that preserve termination, whose iteration is equivalent to strong reduction [28], [54]. We aim to show that such properties derive from generic theorems of higher-order rewriting [110], so that weak reduction can easily be generalised to richer systems with effects.

3.3.4. Direct models and categorical coherence

Proof theory and rewriting are a source of *coherence theorems* in category theory, which show how calculations in a category can be simplified with an embedding into a structure with stronger properties [84], [75]. We aim to explore such results for categorical models of effects [79], [50]. Our key insight is to consider the reflection between *indirect and direct models* [56], [93] as a coherence theorem: it allows us to embed the traditional models of effects into structures for which the rewriting and proof-theoretic techniques from the previous section are effective.

Building on this, we are further interested in connecting operational semantics to 2-category theory, in which a second dimension is traditionally considered for modelling conversions of programs rather than equivalences. This idea has been successfully applied for the λ-calculus [72], [68] but does not scale yet to more realistic models of computation. In our approach, it has already been noticed that the expected symmetries coming from categorical dualities are better represented, motivating a new investigation into this long-standing question.

3.3.5. Models of effects and resources

The unified theory of effects and resources [50] prompts an investigation into the semantics of safe and automatic resource management, in the style of Modern C++ and Rust. Our goal is to show how advanced semantics of effects, resources, and their combination arise by assembling elementary blocks, pursuing the methodology applied by Melliès and Tabareau in the context of continuations [88]. For instance, by combining control flow (exceptions, return) with linearity allows us to describe in a precise way the “Resource Acquisition Is Initialisation” idiom in which the resource safety is ensured with scope-based destructors. A further step would be to reconstruct uniqueness types and borrowing using similar ideas.

3.4. Language extensions for the scaling of proof assistants

The development of tools to construct software systems that respect a given specification is a major challenge of current and future research in computer science. Certified programming with dependent types has recently attracted a lot of interest, and Coq is the *de facto* standard for such endeavours, with an increasing number of users, pedagogical resources, and large-scale projects. Nevertheless, significant work remains to be done to make Coq more usable from a software engineering point of view. The Gallinette team proposes to make progress on three lines of work: (i) the development of gradual certified programming, (ii) the integration of imperative features and object polymorphism in Coq, and (iii) the development of robust tactics for proof engineering for the scaling of formalised libraries.

3.4.1. Gradual Certified Programming

One of the main issues faced by a programmer starting to internalise in a proof assistant code written in a more permissive world is that type theory is constrained by a strict type discipline which lacks flexibility. Concretely, as soon that you start giving more a precise type/specification to a function, the rest of the code interacting with this functions needs to be more precise too. To address this issue, the Gallinette team will put strong efforts into the development of gradual typing in type theory to allow progressive integration of code that comes from a more permissive world.

Indeed, on the way to full verification, programmers can take advantage of a gradual approach in which some properties are simply asserted instead of proven, subject to dynamic verification. Tabareau and Tanter have made preliminary progress in this direction [106]. This work, however, suffers from a number of limitations, the most important being the lack of a mechanism for handling the possibility of runtime errors within Coq. Instead of relying on axioms, this project will explore the application of Section 3.3 to embed effects in Coq. This way, instead of postulating axioms for parts of the development that are too hard/marginal to be dealt with, the system adds dynamic checks. Then, after extraction, we get a program that corresponds to the initial program but with dynamic check for parts that have not been proven, ensuring that the program will raise an error instead of going outside its specification.

This will yield new foundations of gradual certified programming, both more expressive and practical. We will also study how to integrate previous techniques with the extraction mechanism of Coq programs to OCaml, in order to exploit the exception mechanism of OCaml.

3.4.2. Imperative features and object polymorphism in the Coq proof assistant

3.4.2.1. Imperative features.

Abstract data types (ADTs) become useful as the size of programs grows since they provide for a modular approach, allowing abstractions about data to be expressed and then instantiated. Moreover, ADTs are natural concepts in the calculus of inductive constructions. But while it is easy to declare an ADT, it is often difficult to implement an efficient one. Compare this situation with, for example, Okasaki's purely functional data structures [96] which implement ADTs like queues in languages with imperative features. Of course, Okasaki's queues enforce some additional properties for free, such as persistence, but the programmer may prefer to use and to study a simpler implementation without those additional properties. Also in certified symbolic computation (see 3.5.3), an efficient functional implementation of ADTs is often not available, and efficiency is a major challenge in this area. Relying on the theoretical work done in 3.3, we will equip Coq with imperative features and we will demonstrate how they can be used to provide efficient implementations of ADTs. However, it is also often the case that imperative implementation are hard-to-reason-on, requiring for instance the use of separation logic. But in that case, we could take benefice of recent works on integration of separation logic in the Coq proof assistant and in particular the Iris project http://iris-project.org/.

3.4.2.2. Object polymorphism.

Object-oriented programming has evolved since its foundation based on the representation of computations as an exchange of messages between objects. In modern programming languages like Scala, which aims at a synthesis between object-oriented and functional programming, object-orientation concretely results in the use of hierarchies of interfaces ordered by the subtyping relation and the definition of interface implementations that can interoperate. As observed by Cook and Aldrich [48], [31], interoperability can be considered as the essential feature of objects and is a requirement for many modern frameworks and ecosystems: it means that two different implementations of the same interface can interoperate.

Our objective is to provide a representation of object-oriented programs, by focusing on subtyping and interoperability.

For subtyping, the natural solution in type theory is coercive subtyping [82], as implemented in Coq, with an explicit operator for coercions. This should lead to a shallow embedding, but has limitations: indeed, while it allows subtyping to be faithfully represented, it does not provide a direct means to represent union and intersection types, which are often associated with subtyping (for instance intersection types are present in

Scala). A more ambitious solution would be to resort to subsumptive subtyping (or semantic subtyping [55]): in its more general form, a type algebra is extended with boolean operations (union, intersection, complementing) to get a boolean algebra with operators (the original type constructors). Subtying is then interpreted as the natural partial order of the boolean algebra.

We propose to use the type class machinery of Coq to implement semantic subtyping for dependent type theory. Using type class resolution, we can emulate inference rules of subsumptive subtyping without modifying Coq internally. This has also another advantage. As subsumptive subtyping for dependent types should be undecidable in general, using type class resolution allows for an incomplete yet extensible decision procedure.

3.4.3. Robust tactics for proof engineering for the scaling of formalised libraries

When developing certified software, a major part of the effort is spent not only on writing proof scripts, but on *rewriting* them, either for the purpose of code maintenance or because of more significant changes in the base definitions. Regrettably, proof scripts suffer more often than not from a bad programming style, and too many proof developers casually neglect the most elementary principles of well-behaved programmers. As a result, many proof scripts are very brittle, user-defined tactics are often difficult to extend, and sometimes even lack a clear specification. Formal libraries are thus generally very fragile pieces of software. One reason for this unfortunate situation is that proof engineering is very badly served by the tools currently available to the users of the Coq proof assistant, starting with its tactic language. One objective of the Gallinette team is to develop better tools to write proof scripts.

Completing and maintaining a large corpus of formalised mathematics requires a well-designed tactic language. This language should both accommodate the possible specific needs of the theories at stake, and help with diagnostics at refactoring time. Coq's tactic language is in fact two-leveled. First, it includes a basic tactic language, to organise the deductive steps in a proof script and to perform the elementary bureaucracy. Its second layer is a meta-programming language, which allows user to defined their own new tactics at toplevel. Our first direction of work consists in the investigation of the appropriate features of the *basic tactic language*. For instance, the design of the Ssreflect tactic language, and its support for the small scale reflection methodology [61], has been a key ingredient in at least two large scale formalisation endeavours: the Four Colour Theorem [60] and of the Odd Order Theorem [59]. Building on our experience with the Ssreflect tactic language, we will contribute to the ongoing work on the basic tactic language for Coq. The second objective of this task is to contribute to the design of a *typed tactic language*. In particular, we will build on the work of Ziliani and his collaborators [109], extending it with reasoning about the effects that tactics have on the "state of a proof" (e.g. number of sub-goals, metavariables in context). We will also develop a novel approach for incremental type checking of proof scripts, so that programmers gain access to a richer discovery- engineering interaction with the proof assistant.

3.5. Practical experiments

The first three axes of the EPC Gallinette aim at developing a new generation of proof assistants. But we strongly believe that foundational investigations must go hand in hand with practical experiments. Therefore, we expect to benefit from existing expertise and collaborations in the team to experiment our extensions of Coq on real world developments. It should be noticed that those practical experiments are strongly guided by the deep history of research on software engineering of team members.

3.5.1. Certified Code Refactoring

In the context of refactoring of C programs, we intend to formalise program transformations that are written in an imperative style to test the usability of our addition of effects in the proof assistant. This subject has been chosen based on the competence of members of the team.

We are currently working on the formalisation of refactoring tools in Coq [44]. Automatic refactoring of programs in industrial languages is difficult because of the large number of potential interactions between language features that are difficult to predict and to test. Indeed, all available refactoring tools suffer from bugs : they fail to ensure that the generated program has the same behaviour as the input program. To cope with that difficulty, we have chosen to build a refactoring tool with Coq : a program transformation is written in the Coq programming language, then proven correct on all possible inputs, and then an OCaml executable program is generated by the platform. We rely on the CompCert C formalisation of the C language. CompCert is currently the most complete formalisation of an industrial language, which justifies that choice. We have three goals in that project :

- Build a refactoring tool that programmers can rely on and make it available in a popular platform (such as Eclipse, IntelliJ or Frama-C).
- Explore large, drastic program transformations such as replacing a design architecture for an other one, by applying a sequence of small refactoring operations (as we have done for Java and Haskell programs before [47], [43], [30]), while ensuring behaviour preservation.
- Explore the use of enhancements of proof systems on large developments. For instance, refactoring tools are usually developed in the imperative/object paradigm, so the extension of Coq with side effects or with object features proposed in the team can find a direct use-case here.

3.5.2. Certified Constraint Programming

We plan to make use of the internalisation of the object-oriented paradigm in the context of constraint programming. Indeed, this domain is made of very complex algorithms that are often developed using object-oriented programming (as it is the case for instance for CHOCO, which is developed in the Tasc Group at IMT Atlantique, Nantes). We will in particular focus on filtering algorithms in constraint solvers, for which research publications currently propose new algorithms with manual proofs. Their formalisation in Coq is challenging. Another interesting part of constraint solving to formalise is the part that deals with program generation (as opposed to extraction). However, when there are numerous generated pieces of code, it is not realistic to prove their correctness manually, and it can be too difficult to prove the correctness of a generator. So we intend to explore a middle path that consists in generating a piece of code along with its corresponding proof (script or proof term). A target application could be interval constraints (for instance Allen interval algebra or region connection calculus) that can generate thousands of specialised filtering algorithms for a small number of variables [36].

Finally, Rémi Douence has already worked (articles publishing [63], [97], [53], PhD Thesis advising [98]) with different members of the Tasc team. Currently, he supervises with Nicolas Beldiceanu the PhD Thesis of Ekaterina Arafailova in the Tasc team. She studies finite transducers to model time-series constraints [37], [35], [34]. This work requires proofs, manually done for now, we would like to explore when these proofs could be mechanised.

3.5.3. Certified Symbolic Computation

We will investigate how the addition of effects in the Coq proof assistant can facilitate the marriage of computer algebra with formal proofs. Computer algebra systems on one hand, and proof assistants on the other hand, are both designed for doing mathematics with the help of a computer, by the means of symbolic computations. These two families of systems are however very different in nature: computer algebra systems allow for implementations faithful to the theoretical complexity of the algorithms, whereas proof assistants have the expressiveness to specify exactly the semantic of the data-structures and computations.

Experiments have been run that link computer algebra systems with Coq [52], [42]. These bridges rely on the implementation of formal proof-producing core algorithms like normalisation procedures. Incidentally, they require non trivial maintenance work to survive the evolution of both systems. Other proof assistants like the Isabelle/HOL system make use of so-called reflection schemes: the proof assistant can produce code in an external programming language like SML, but also allows to import the values output by these extracted programs back inside the formal proofs. This feature extends the trusted base of code quite significantly but it has been used for major achievements like a certified symbolic/numeric ODE solver [69].

We would like to bring Coq closer to the efficiency and user-friendliness of computer algebra systems: for now it is difficult to use the Coq programming language so that certified implementations of computer algebra algorithms have the right, observable, complexity when they are executed inside Coq. We see the addition of effects to the proof assistant as an opportunity to ease these implementations, for instance by making use of caching mechanisms or of profiling facilities. Such enhancements should enable the verification of computation-intensive mathematical proofs that are currently beyond reach, like the validation of Helfgott's proof of the weak Goldbach conjecture [65].

4. Highlights of the Year

4.1. Highlights of the Year

4.1.1. Permanents members

Gaëtan Gilbert, currently PhD student in the Gallinette team, will be promoted expert engineer for the Coq consortium, staying in the Gallinette team.

Matthieu Sozeau, Inria Junior Researcher and leader of the Coq development team, is joining the Gallinette team end of 2019-beginning of 2020.

Nicolas Tabareau is now director of research (DR2) at Inria since October 2019.

4.1.2. Awards

Marie Kerjean has been awarded a L'Oréal - Unesco Foundation grant.

L'Oréal - Unesco Grants for Women in Science are awarded to talented young female researchers.

5. New Software and Platforms

5.1. Coq

The Coq Proof Assistant

KEYWORDS: Proof - Certification - Formalisation

SCIENTIFIC DESCRIPTION: Coq is an interactive proof assistant based on the Calculus of (Co-)Inductive Constructions, extended with universe polymorphism. This type theory features inductive and co-inductive families, an impredicative sort and a hierarchy of predicative universes, making it a very expressive logic. The calculus allows to formalize both general mathematics and computer programs, ranging from theories of finite structures to abstract algebra and categories to programming language metatheory and compiler verification. Coq is organised as a (relatively small) kernel including efficient conversion tests on which are built a set of higher-level layers: a powerful proof engine and unification algorithm, various tactics/decision procedures, a transactional document model and, at the very top an IDE.

FUNCTIONAL DESCRIPTION: Coq provides both a dependently-typed functional programming language and a logical formalism, which, altogether, support the formalisation of mathematical theories and the specification and certification of properties of programs. Coq also provides a large and extensible set of automatic or semi-automatic proof methods. Coq's programs are extractible to OCaml, Haskell, Scheme, ...

RELEASE FUNCTIONAL DESCRIPTION: Coq version 8.10 contains two major new features: support for a native fixed-precision integer type and a new sort SProp of strict propositions. It is also the result of refinements and stabilization of previous features, deprecations or removals of deprecated features, cleanups of the internals of the system and API, and many documentation improvements. This release includes many user-visible changes, including deprecations that are documented in the next subsection, and new features that are documented in the reference manual.

Version 8.10 is the fifth release of Coq developed on a time-based development cycle. Its development spanned 6 months from the release of Coq 8.9. Vincent Laporte is the release manager and maintainer of this release. This release is the result of 2500 commits and 650 PRs merged, closing 150+ issues.

See the Zenodo citation for more information on this release: https://zenodo.org/record/3476303#.Xe54f5NKjOQ

NEWS OF THE YEAR: Coq 8.10.0 contains:

- some quality-of-life bug fixes, - a critical bug fix related to template polymorphism, - native 63-bit machine integers, - a new sort of definitionally proof-irrelevant propositions: SProp, - private universes for opaque polymorphic constants, - string notations and numeral notations, - a new simplex-based proof engine for the tactics lia, nia, lra and nra, - new introduction patterns for SSReflect, - a tactic to rewrite under binders: under, - easy input of non-ASCII symbols in CoqIDE, which now uses GTK3.

All details can be found in the user manual.

- Participants: Yves Bertot, Frédéric Besson, Maxime Denes, Emilio Jesús Gallego Arias, Gaëtan Gilbert, Jason Gross, Hugo Herbelin, Assia Mahboubi, Érik Martin-Dorel, Guillaume Melquiond, Pierre-Marie Pédrot, Michael Soegtrop, Matthieu Sozeau, Enrico Tassi, Laurent Théry, Théo Zimmermann, Theo Winterhalter, Vincent Laporte, Arthur Charguéraud, Cyril Cohen, Christian Doczkal and Chantal Keller
- Partners: CNRS - Université Paris-Sud - ENS Lyon - Université Paris-Diderot
- Contact: Matthieu Sozeau
- URL: http://coq.inria.fr/

5.2. Math-Components

Mathematical Components library

KEYWORD: Proof assistant

FUNCTIONAL DESCRIPTION: The Mathematical Components library is a set of Coq libraries that cover the prerequiste for the mechanization of the proof of the Odd Order Theorem.

RELEASE FUNCTIONAL DESCRIPTION: The library includes 16 more theory files, covering in particular field and Galois theory, advanced character theory, and a construction of algebraic numbers.

- Participants: Alexey Solovyev, Andrea Asperti, Assia Mahboubi, Cyril Cohen, Enrico Tassi, François Garillot, Georges Gonthier, Ioana Pasca, Jeremy Avigad, Laurence Rideau, Laurent Théry, Russell O'Connor, Sidi Ould Biha, Stéphane Le Roux and Yves Bertot
- Contact: Assia Mahboubi
- URL: http://math-comp.github.io/math-comp/

5.3. Ssreflect

FUNCTIONAL DESCRIPTION: Ssreflect is a tactic language extension to the Coq system, developed by the Mathematical Components team.

- Participants: Assia Mahboubi, Cyril Cohen, Enrico Tassi, Georges Gonthier, Laurence Rideau, Laurent Théry and Yves Bertot
- Contact: Yves Bertot
- URL: http://math-comp.github.io/math-comp/

5.4. Ltac2

KEYWORDS: Coq - Proof assistant

FUNCTIONAL DESCRIPTION: A replacement for Ltac, the tactic language of Coq.

- Contact: Pierre-Marie Pédrot

6. New Results

6.1. Logical Foundations of Programming Languages

Participants: Esaie Bauer, Rémi Douence, Marie Kerjean, Ambroise Lafont, Maxime Lucas, Étienne Miquey, Guillaume Munch-Maccagnoni, Nicolas Tabareau.

6.1.1. Classical Logic

6.1.1.1. Continuation-and-environment-passing style translations: a focus on call-by-need

The call-by-need evaluation strategy for the λ-calculus is an evaluation strategy that lazily evaluates arguments only if needed, and if so, shares computations across all places where it is needed. To implement this evaluation strategy, abstract machines require some form of global environment. While abstract machines usually lead to a better understanding of the flow of control during the execution, easing in particular the definition of continuation-passing style translations, the case of machines with global environments turns out to be much more subtle. The main purpose of [21] is to understand how to type a continuation-and-environment-passing style translations, that it to say how to soundly translate a classical calculus with environment into a calculus that does not have these features. To this end, we focus on a sequent calculus presentation of a call-by-need λ-calculus with classical control for which Ariola et. al already defined an untyped translation and which we equipped with a system of simple types in a previous paper. We present here a type system for the target language of their translation, which highlights a variant of Kripke forcing related to the environment-passing part of the translation. Finally, we show that our construction naturally handles the cases of call-by-name and call-by-value calculi with environment, encompassing in particular the Milner Abstract Machine, a machine with global environments for the call-by-name λ-calculus.

6.1.1.2. Revisiting the duality of computation: an algebraic analysis of classical realizability models

In an impressive series of papers, Krivine showed at the edge of the last decade how classical realizability provides a surprising technique to build models for classical theories. In particular, he proved that classical realizability subsumes Cohen's forcing, and even more, gives rise to unexpected models of set theories. Pursuing the algebraic analysis of these models that was first undertaken by Streicher, Miquel recently proposed to lay the algebraic foundation of classical realizability and forcing within new structures which he called implicative algebras. These structures are a generalization of Boolean algebras based on an internal law representing the implication. Notably, implicative algebras allow for the adequate interpretation of both programs (i.e. proofs) and their types (i.e. formulas) in the same structure. The very definition of implicative algebras takes position on a presentation of logic through universal quantification and the implication and, computationally, relies on the call-by-name λ-calculus. In [13], we investigate the relevance of this choice, by introducing two similar structures. On the one hand, we define disjunctive algebras, which rely on internal laws for the negation and the disjunction and which we show to be particular cases of implicative algebras. On the other hand, we introduce conjunctive algebras, which rather put the focus on conjunctions and on the call-by-value evaluation strategy. We finally show how disjunctive and conjunctive algebras algebraically reflect the well-known duality of computation between call-by-name and call-by-value.

6.1.2. Models of programming languages mixing effects and resources

6.1.2.1. Efficient deconstruction with typed pointer reversal

Building on the connection between resource management in systems programming and ordered logic we established previously, we investigate a pervasive issue in the languages C++ and Rust whereby compiler-generated clean-up functions cause a stack overflow on deep structures. In [17], we show how to generate clean-up algorithms that run in constant time and space for a broad class of ordered algebraic datatypes such as ones that can be found in C++ and Rust or in future extensions of functional programming languages with first-class resources.

6.1.2.2. Resource safety in OCaml

Building on our investigations for a resource-management model for OCaml, we have proposed several preliminary improvements to the OCaml language. We contributed to the design and implementation of new resource management primitives (PRs #2118, #8962), resource-safe C APIs (PRs #8993, #8997, #9037), and core runtime capabilities (PR #8961). (#2118 has been merged into OCaml 4.08 and #8993 and #9037 have been merged into OCaml 4.10.)

We continued to interact with L. White and S. Dolan (Jane Street), on the design of resource management and exception safety in multicore OCaml.

6.1.3. Syntax and Rewriting Systems

6.1.3.1. Reduction Monads and Their Signatures

In [1], we study reduction monads, which are essentially the same as monads relative to the free functor from sets into multigraphs. Reduction monads account for two aspects of the lambda calculus: on the one hand, in the monadic viewpoint, the lambda calculus is an object equipped with a well-behaved substitution; on the other hand, in the graphical viewpoint, it is an oriented multigraph whose vertices are terms and whose edges witness the reductions between two terms. We study presentations of reduction monads. To this end, we propose a notion of reduction signature. As usual, such a signature plays the role of a virtual presentation, and specifies arities for generating operations-possibly subject to equations-together with arities for generating reduction rules. For each such signature, we define a category of models; any model is, in particular, a reduction monad. If the initial object of this category of models exists, we call it the reduction monad presented (or specified) by the given reduction signature. Our main result identifies a class of reduction signatures which specify a reduction monad in the above sense. We show in the examples that our approach covers several standard variants of the lambda calculus.

6.1.3.2. Modules over monads and operational semantics

[22] is a contribution to the search for efficient and high-level mathematical tools to specify and reason about (abstract) programming languages or calculi. Generalising the reduction monads of Ahrens et al., we introduce operational monads, thus covering new applications such as the-calculus, Positive GSOS specifications, and the big-step, simply-typed, call-by-value-calculus. Finally, we design a notion of signature for operational monads that covers all our examples.

6.1.3.3. Modular specification of monads through higher-order presentations

In their work on second-order equational logic, Fiore and Hur have studied presentations of simply typed languages by generating binding constructions and equations among them. To each pair consisting of a binding signature and a set of equations, they associate a category of 'models', and they give a monadicity result which implies that this category has an initial object, which is the language presented by the pair. In [10], we propose, for the untyped setting, a variant of their approach where monads and modules over them are the central notions. More precisely, we study, for monads over sets, presentations by generating ('higher-order') operations and equations among them. We consider a notion of 2-signature which allows to specify a monad with a family of binding operations subject to a family of equations, as is the case for the paradigmatic example of the lambda calculus, specified by its two standard constructions (application and abstraction) subject to β- and η-equalities. Such a 2-signature is hence a pair (Σ,E) of a binding signature Σ and a family E of equations for Σ. This notion of 2-signature has been introduced earlier by Ahrens in a slightly different context. We associate, to each 2-signature (Σ,E), a category of 'models of (Σ,E); and we say that a 2-signature is 'effective' if this category has an initial object; the monad underlying this (essentially unique) object is the 'monad specified by the 2-signature'. Not every 2-signature is effective; we identify a class of 2-signatures, which we call 'algebraic', that are effective. Importantly, our 2-signatures together with their models enjoy 'modularity': when we glue (algebraic) 2-signatures together, their initial models are glued accordingly. We provide a computer formalization for our main results.

6.1.3.4. The Diamond Lemma for non-terminating rewriting systems

In [16], we study the confluence property for rewriting systems whose underlying set of terms admits a vector space structure. For that, we use deterministic reduction strategies. These strategies are based on the choice of standard reductions applied to basis elements. We provide a sufficient condition of confluence in terms of the kernel of the operator which computes standard normal forms. We present a local criterion which enables us to check the confluence property in this framework. We show how this criterion is related to the Diamond Lemma for terminating rewriting systems

6.1.4. Differential Linear Logic

6.1.4.1. Higher-order distributions for differential linear logic

Linear Logic was introduced as the computational counterpart of the algebraic notion of linearity. Differential Linear Logic refines Linear Logic with a proof-theoretical interpretation of the geometrical process of differentiation. In [24], we construct a polarized model of Differential Linear Logic satisfying computational constraints such as an interpretation for higher-order functions, as well as constraints inherited from physics such as a continuous interpretation for spaces. This extends what was done previously by Kerjean for first order Differential Linear Logic without promotion. Concretely, we follow the previous idea of interpreting the exponential of Differential Linear Logic as a space of higher-order distributions with compact-support, and is constructed as an inductive limit of spaces of distributions on Euclidean spaces. We prove that this exponential is endowed with a co-monadic like structure, with the notable exception that it is functorial only on isomorphisms. Interestingly, as previously argued by Ehrhard, this still allows one to interpret differential linear logic without promotion.

6.1.4.2. Chiralities in topological vector spaces

Chiralities are categories introduced by Mellies to account for a game semantics point of view on negation. In [23], [20], we uncover instances of this structure in the theory of topological vector spaces, thus constructing several new polarized models of Multiplicative Linear Logic. These models improve previously known smooth models of Differential Linear Logic, showing the relevance of chiralities to express topological properties of vector spaces. They are the first denotational polarized models of Multiplicative Linear Logic, based on the pre-existing theory of topological vector spaces, in which two distinct sets of formulas, two distinct negations, and two shifts appear naturally.

6.1.5. Distributed Programming

6.1.5.1. Chemical foundations of distributed aspects.

Distributed applications are challenging to program because they have to deal with a plethora of concerns, including synchronisation, locality, replication, security and fault tolerance. Aspect-oriented programming (AOP) is a paradigm that promotes better modularity by providing means to encapsulate cross-cutting concerns in entities called aspects. Over the last years, a number of distributed aspect-oriented programming languages and systems have been proposed, illustrating the benefits of AOP in a distributed setting. Chemical calculi are particularly well-suited to formally specify the behaviour of concurrent and distributed systems. The join calculus is a functional name-passing calculus, with both distributed and object-oriented extensions. It is used as the basis of concurrency and distribution features in several mainstream languages like C# (Polyphonic C#, now Cω), OCaml (JoCaml), and Scala Joins. Unsurprisingly, practical programming in the join calculus also suffers from modularity issues when dealing with crosscutting concerns. We propose the Aspect Join Calculus [9], an aspect-oriented and distributed variant of the join calculus that addresses crosscutting and provides a formal foundation for distributed AOP. We develop a minimal aspect join calculus that allows aspects to advise chemical reactions. We show how to deal with causal relations in pointcuts and how to support advanced customisable aspect weaving semantics.

6.2. Type Theory and Proof Assistants

Participants: Simon Boulier, Gaëtan Gilbert, Maxime Lucas, Pierre-Marie Pédrot, Loïc Pujet, Nicolas Tabareau, Théo Winterhalter.

6.2.1. Type Theory

6.2.1.1. Effects in Type Theory.

There is a critical tension between substitution, dependent elimination and effects in type theory. In this paper, we crystallize this tension in the form of a no-go theorem that constitutes the fire triangle of type theory. To release this tension, we propose in [7] DCBPV, an extension of call-by-push-value (CBPV)-a general calculus of effects-to dependent types. Then, by extending to CBPV the well-known decompositions of call-by-name and call-by-value into CBPV, we show why, in presence of effects, dependent elimination must be restricted in call-by-name, and substitution must be restricted in call-by-value. To justify DCBPV and show that it is general enough to interpret many kinds of effects, we define various effectful syntactic translations from DCBPV to Martin-Löf type theory: the reader, weaning and forcing translations.

Traditional approaches to compensate for the lack of exceptions in type theories for proof assistants have severe drawbacks from both a programming and a reasoning perspective. We recently extended the Calculus of Inductive Constructions (CIC) with exceptions. The new exceptional type theory is interpreted by a translation into CIC, covering full dependent elimination, decidable type-checking and canonicity. However, the exceptional theory is inconsistent as a logical system. To recover consistency, we propose an additional translation that uses parametricity to enforce that all exceptions are caught locally. While this enforcement brings logical expressivity gains over CIC, it completely prevents reasoning about exceptional programs such as partial functions. In [6], we addresses the dilemma between exceptions and consistency in a more flexible manner, with the Reasonably Exceptional Type Theory (RETT). RETT is structured in three layers: (a) the exceptional layer, in which all terms can raise exceptions; (b) the mediation layer, in which exceptional terms must be provably parametric; (c) the pure layer, in which terms are non-exceptional, but can refer to exceptional terms. We present the general theory of RETT, where each layer is realized by a predicative hierarchy of universes, and develop an instance of RETT in Coq: the impure layer corresponds to the predicative universe hierarchy, the pure layer is realized by the impredicative universe of propositions, and the mediation layer is reified via a parametricity type class. RETT is the first full dependent type theory to support consistent reasoning about exceptional terms, and the CoqRETT plugin readily brings this ability to Coq programmers.

6.2.1.2. Eliminating Reflection from Type Theory.

Type theories with equality reflection, such as extensional type theory (ETT), are convenient theories in which to formalise mathematics, as they make it possible to consider provably equal terms as convertible. Although type-checking is undecidable in this context, variants of ETT have been implemented, for example in NuPRL and more recently in Andromeda. The actual objects that can be checked are not proof-terms, but derivations of proof-terms. This suggests that any derivation of ETT can be translated into a typecheckable proof term of intensional type theory (ITT). However, this result, investigated categorically by Hofmann in 1995, and 10 years later more syntactically by Oury, has never given rise to an effective translation. In [15], we provide the first syntactical translation from ETT to ITT with uniqueness of identity proofs and functional extensionality. This translation has been defined and proven correct in Coq and yields an executable plugin that translates a derivation in ETT into an actual Coq typing judgment. Additionally, we show how this result is extended in the context of homotopy to a two-level type theory.

6.2.1.3. Setoid type theory - a syntactic translation

[11] introduces setoid type theory, an intensional type theory with a proof-irrelevant universe of propositions and an equality type satisfying function extensionality, propositional extensionality and a definitional computation rule for transport. We justify the rules of setoid type theory by a syntactic translation into a pure type theory with a universe of propositions. We conjecture that our syntax is complete with regards to this translation.

6.2.1.4. The folk model category structure on strict ω-categories is monoidal

In [19], we prove that the folk model category structure on the category of strict ω-categories, introduced by Lafont, Métayer and Worytkiewicz, is monoidal, first, for the Gray tensor product and, second, for the join of ω-categories, introduced by the first author and Maltsiniotis. We moreover show that the Gray tensor product induces, by adjunction, a tensor product of strict (m, n)-categories and that this tensor product is also compatible with the folk model category structure. In particular, we get a monoidal model category structure on the category of strict ω-groupoids. We prove that this monoidal model category structure satisfies the monoid axiom, so that the category of Gray monoids, studied by the second author, bears a natural model category structure.

6.2.2. Proof Assistants

6.2.2.1. Metacoq

The MetaCoq project [26], [26] aims to provide a certified meta-programming environment in Coq. It builds on Template-Coq, a plugin for Coq originally implemented by Malecha (2014), which provided a reifier for Coq terms and global declarations, as represented in the Coq kernel, as well as a denotation command. Recently, it was used in the CertiCoq certified compiler project (Anand et al., 2017), as its front-end language, to derive parametricity properties (Anand and Morrisett, 2018). However, the syntax lacked semantics, be it typing semantics or operational semantics, which should reflect, as formal specifications in Coq, the semantics of Coq's type theory itself. The tool was also rather bare bones, providing only rudimentary quoting and unquoting commands. We generalize it to handle the entire Polymorphic Calculus of Cumulative Inductive Constructions (pCUIC), as implemented by Coq, including the kernel's declaration structures for definitions and inductives, and implement a monad for general manipulation of Coq's logical environment. We demonstrate how this setup allows Coq users to define many kinds of general purpose plugins, whose correctness can be readily proved in the system itself, and that can be run efficiently after extraction. We give a few examples of implemented plugins, including a parametricity translation and a certifying extraction to call-by-value λ-calculus. We also advocate the use of MetaCoq as a foundation for higher-level tools.

6.2.2.2. Verification of Type Checking and Erasure for Coq, in Coq

Coq is built around a well-delimited kernel that perfoms typechecking for definitions in a variant of the Calculus of Inductive Constructions (CIC). Although the metatheory of CIC is very stable and reliable, the correctness of its implementation in Coq is less clear. Indeed, implementing an efficient type checker for CIC is a rather complex task, and many parts of the code rely on implicit invariants which can easily be broken by further evolution of the code. Therefore, on average, one critical bug has been found every year in Coq. [8] presents the first implementation of a type checker for the kernel of Coq (without the module system and template polymorphism), which is proven correct in Coq with respect to its formal specification and axiomatisation of part of its metatheory. Note that because of Gödel's incompleteness theorem, there is no hope to prove completely the correctness of the specification of Coq inside Coq (in particular strong normalisation or canonicity), but it is possible to prove the correctness of the implementation assuming the correctness of the specification, thus moving from a trusted code base (TCB) to a trusted theory base (TTB) paradigm. Our work is based on the MetaCoq project which provides metaprogramming facilities to work with terms and declarations at the level of this kernel. Our type checker is based on the specification of the typing relation of the Polymorphic, Cumulative Calculus of Inductive Constructions (pCUIC) at the basis of Coq and the verification of a relatively efficient and sound type-checker for it. In addition to the kernel implementation, an essential feature of Coq is the so-called extraction: the production of executable code in functional languages from Coq definitions. We present a verified version of this subtle type-and-proof erasure step, therefore enabling the verified extraction of a safe type-checker for Coq.

6.2.2.3. Definitional Proof-Irrelevance without K.

Definitional equality—or conversion—for a type theory with a decidable type checking is the simplest tool to prove that two objects are the same, letting the system decide just using computation. Therefore, the more things are equal by conversion, the simpler it is to use a language based on type theory. Proof-irrelevance, stating that any two proofs of the same proposition are equal, is a possible way to extend conversion to make a type theory more powerful. However, this new power comes at a price if we integrate it naively, either by making type checking undecidable or by realising new axioms—such as uniqueness of identity proofs (UIP)—that are incompatible with other extensions, such as univalence. In [3], taking inspiration from homotopy type theory, we propose a general way to extend a type theory with definitional proof irrelevance, in a way that keeps type checking decidable and is compatible with univalence. We provide a new criterion to decide whether a proposition can be eliminated over a type (correcting and improving the so-called singleton elimination of Coq) by using techniques coming from recent development on dependent pattern matching without UIP. We show the generality of our approach by providing implementations for both Coq and Agda, both of which are planned to be integrated in future versions of those proof assistants.

6.2.2.4. Cubical Synthetic Homotopy Theory

Homotopy type theory is an extension of type theory that enables synthetic reasoning about spaces and homotopy theory. This has led to elegant computer formalizations of multiple classical results from homotopy theory. However, many proofs are still surprisingly complicated to formalize. One reason for this is the axiomatic treatment of univalence and higher inductive types which complicates synthetic reasoning as many intermediate steps, that could hold simply by computation, require explicit arguments. Cubical type theory offers a solution to this in the form of a new type theory with native support for both univalence and higher inductive types. In [14], we show how the recent cubical extension of Agda can be used to formalize some of the major results of homotopy type theory in a direct and elegant manner.

6.3. Program Certifications and Formalisation of Mathematics

Participants: Julien Cohen, Rémi Douence, Guilhem Jaber, Assia Mahboubi, Igor Zhirkov.

6.3.1. CoqTL: A Coq DSL for Rule-Based Model Transformation

In model-driven engineering, model transformation (MT) verification is essential for reliably producing software artifacts. While recent advancements have enabled automatic Hoare-style verification for non-trivial MTs, there are certain verification tasks (e.g. induction) that are intrinsically difficult to automate. Existing tools that aim at simplifying the interactive verification of MTs typically translate the MT specification (e.g. in ATL) and properties to prove (e.g. in OCL) into an interactive theorem prover. However, since the MT specification and proof phases happen in separate languages, the proof developer needs a detailed knowledge of the translation logic. Naturally, any error in the MT translation could cause unsound verification, i.e. the MT executed in the original environment may have different semantics from the verified MT. In [2] , we propose an alternative solution by designing and implementing an internal domain specific language, namely CoqTL, for the specification of declarative MTs directly in the Coq interactive theorem prover. Expressions in CoqTL are written in Gallina (the specification language of Coq), increasing the possibilities of reusing native Coq libraries in the transformation definition and proof. CoqTL specifications can be directly executed by our transformation engine encoded in Coq, or a certified implementation of the transformation can be generated by the native Coq extraction mechanism. We ensure that CoqTL has the same expressive power of Gallina (i.e. if a MT can be computed in Gallina, then it can also be represented in CoqTL). In this article, we introduce CoqTL, evaluate its practical applicability on a use case, and identify its current limitations.

6.3.2. A certificate-based approach to formally verified approximations.

In [12], we present a library to verify rigorous approximations of univariate functions on real numbers, with the Coq proof assistant. Based on interval arithmetic, this library also implements a technique of validation a posteriori based on the Banach fixed-point theorem. We illustrate this technique on the case of operations of division and square root. This library features a collection of abstract structures that organise the specification of rigorous approximations, and modularise the related proofs. Finally, we provide an implementation of verified Chebyshev approximations, and we discuss a few examples of computations.

6.3.3. Formally Verified Approximations of Definite Integrals.

Finding an elementary form for an antiderivative is often a difficult task, so numerical integration has become a common tool when it comes to making sense of a definite integral. Some of the numerical integration methods can even be made rigorous: not only do they compute an approximation of the integral value but they also bound its inaccuracy. Yet numerical integration is still missing from the toolbox when performing formal proofs in analysis. In [5], we present an efficient method for automatically computing and proving bounds on some definite integrals inside the Coq formal system. Our approach is not based on traditional quadrature methods such as Newton-Cotes formulas. Instead, it relies on computing and evaluating antiderivatives of rigorous polynomial approximations, combined with an adaptive domain splitting. Our approach also handles improper integrals, provided that a factor of the integrand belongs to a catalog of identified integrable functions. This work has been integrated to the CoqInterval library.

6.3.4. Reasoning about exact memory transformations induced by refactorings in CompCert C

[18] reports on our work in extending CompCert memory model with a relation to model relocations. It preserves undefined values unlike similar relations defined in CompCert. This relation commutes with memory operations. Our main contributions are the relation itself and mechanically checked proofs of its commutation properties. We intend to use this extension to construct and verify a refactoring tool for programs written in C.

6.3.5. Automating Contextual Equivalence for Higher-Order Programs with References

In [4], we have proposed a framework to study contextual equivalence of programs written in a call-by-value functional language with local integer references. It reduces the problem of contextual equivalence to the problem of non-reachability in a transition system of memory configurations. This reduction is complete for recursion-free programs. Restricting to programs that do not allocate references inside the body of functions, we have encoded this non-reachability problem as a set of constrained Horn clause that can then be checked for satisfiability automatically. Restricting furthermore to a language with finite data-types, we also get a new decidability result for contextual equivalence at any type.

7. Partnerships and Cooperations

7.1. Regional Initiatives

Vercoma (Atlanstic 2020/Attractivity grant)

Goal: Verified computer mathematics.

Coordinator: A. Mahboubi.

Duration: 08/2018 - 08/2021.

7.2. National Initiatives

7.2.1. ANR

FastRelax (ANR-14-CE25-0018).

Goal: Develop computer-aided proofs of numerical values, with certified and reasonably tight error bounds, without sacrificing efficiency.

Coordinator: Bruno Salvy (Inria, ENS Lyon).

Participant: A. Mahboubi.

Duration: 2014-2019.

Website: http://fastrelax.gforge.inria.fr/.

Note: This project started when A. Mahboubi was still in the Specfun project at the Saclay Île-de-France CRI. The budget is still managed there, within the Toccata project, but remains available to A. Mahboubi.

7.3. European Initiatives

7.3.1. FP7 & H2020 Projects

7.3.1.1. 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.

Program: COST

Project acronym: EUTYPES

Project title: The European research network on types for programming and verification

Duration: 21/03/2016 - 20/03/2020.

Coordinator: Herman Geuvers (Radboud Univerity, Nijmegen, The Netherlands)

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.

Europe has a strong type theory community, ranging from foundational research to applications in programming languages, verification and theorem proving, which is in urgent need of better networking. A COST Action that crosses the borders will support the collaboration between groups and complementary expertise, and mobilise a critical mass of existing type theory research.

7.4. International Initiatives

7.4.1. Inria International Labs

Inria Chile

Associate Team involved in the International Lab:

7.4.1.1. GECO

Title: Gradual verification and robust proof Engineering for COq

International Partner (Institution - Laboratory - Researcher):

Universidad de Chile (Chile) - Centrum Wiskunde & Informatica - Éric Tanter

Start year: 2018

See also: http://geco.gforge.inria.fr

The development of tools to construct software systems that respect a given specification is a major challenge of current and future research in computer science. Interactive theorem provers based on type theory, such as Coq, have shown their effectiveness to prove correctness of important pieces of software like the C compiler of the CompCert project. Certified programming with dependent types is attracting a lot of attention recently, and Coq is the de facto standard for such endeavors, with an increasing amount of users, pedagogical material, and large-scale projects. Nevertheless, significant work remains to be done to make Coq more usable from a software engineering point of view.

This collaboration project gathers the expertise of researchers from Chile (Inria Chile, Universidad de Chile, Universidad Católica de Valparaíso) and France (Inria Nantes, Inria Paris), in different areas that are crucial to develop the vision of certified software engineering. The focus of this project is both theoretical and practical, covering novel foundations and methods, design of concrete languages and tools, and validation through specific case studies.

The end result will be a number of enhancements to the Coq proof assistant (frameworks, tactic language) together with guidelines and demonstrations of their applicability in realistic scenarios.

7.4.2. Inria International Partners

7.4.2.1. Informal International Partners

- A. Mahboubi holds a part-time endowed professor position in the Department of Mathematics at the Vrije Universiteit Amsterdam (the Netherlands).

7.5. International Research Visitors

7.5.1. Visits of International Scientists

- Matias Toro (U. Chile) visited 1 week in January to work with G. Munch-Maccagnoni.

7.5.2. Visits to International Teams

7.5.2.1. Research Stays Abroad

+ G. Munch-Maccagnoni visited E. Tanter and M. Toro (U. Chile) in March.

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Scientific Events: Selection

8.1.1.1. Chair of Conference Program Committees

- A. Mahboubi has served as co-PC chair of the 8th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP'19).
- A. Mahboubi has served as co-chair of the CoqPL'20 workshop, satellite of POPL'20.

8.1.1.2. Member of the Conference Program Committees

- P.-M. Pédrot has been a member of the program committee of the Coq Workshop'19 and JFLA'20.
- N. Tabareau has been a member of the program committee of FSCD'19 and POPL'19.
- G. Munch-Maccagnoni has been a member of the external review committee for ICFP'19 conference, and a member of the program committees for the workshops SD'19 (affiliated with FSCD) and LOLA'19 (affilitated with LICS).
- A. Mahboubi has been a member of the program commitee of the ITP'19, Frocos'19 and CPP'20 international conferences, and of the TFP'19 workshop.
- G. Jaber has been a member of the program committee of the Student Research Competition of POPL'20.

8.1.1.3. Reviewer

- P.-M. Pédrot has served as an external reviewer for CPP'19, FoSSaCS'19, ICFP'19 and LICS'19.
- N. Tabareau has served as an external reviewer for LICS'19, CPP'19.
- G. Munch-Maccagnoni has served as an external reviewer for FoSSaCS'20.
- G. Jaber has served as an external reviewer for CONCUR'19, ITP'19, FoSSaCS'19 and POPL'20.

8.1.2. Journal

8.1.2.1. Member of the Editorial Boards

- A. Mahboubi is a member of the editorial board of the Journal of Automated Reasoning.
- A. Mahboubi is co-editing the post-proceedings of the TYPES 2019 conference.

8.1.2.2. Reviewer - Reviewing Activities

- N. Tabareau has been a reviewer for Mathematical Strucutre in Computer Science.

8.1.3. Invited Talks

- P.-M. Pédrot gave a 3 hour long invited class on effectful types theories at the JFLA 2019.
- N. Tabareau has given an invited talk at Coq Workshop 2019 in Portland, Oregon.
- G. Munch-Maccagnoni gave a talk at ITU Copenhaguen in July and a talk at the seminar of philosophy in computer science Codes Sources in Paris in June.
- A. Mahboubi has given an invited talk repectively at the MPC'19, CiE'19, Cade'19 and Types'19 international conferences, at the PxTP'19 international workshop and at the Codes Sources seminar.

8.1.4. Leadership within the Scientific Community

- N. Tabareau is a member of the scientific committee of the GdR of Algebraic Topology.
- A. Mahboubi is a member of the core managment group of the EUTypes project, and leader of the "Tools" working group.
- A. Mahboubi is a member of the steering committee of the ITP conference.
- A. Mahboubi is a member of the scientific committee of the GdR "Informatique Mathématique".

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Licence : Julien Cohen, Discrete Mathematics, 48h, L1 (IUT), IUT Nantes, France

Licence : Julien Cohen, Introduction to proof assistants (Coq), 8h, L2 (PEIP : IUT/Engineering school), Polytech Nantes, France

Licence : Julien Cohen, Functional Programming (Scala), 22h, L2 (IUT), IUT Nantes, France

Master : Julien Cohen, Object oriented programming (Java), 32h, M1 (Engineering school), Polytech Nantes, France

Master : Julien Cohen, Functional programming (OCaml), 18h, M1 (Engineering school), Polytech Nantes, France

Master : Julien Cohen, Tools for softaware engineering (proof with Frama-C, test, code management), 20h, M1 (Engineering school), Polytech Nantes, France

Licence : Rémi Douence, Object Oriented Design and Programming, 45h, L1 (engineers), IMT-Atlantique, Nantes, France

Licence : Rémi Douence, Introduction to scientific research in computer science (Project: an Haskell interpreter in Java) Project, 20h, L1 (engineers), IMT-Atlantique, Nantes, France

Licence : Rémi Douence, Object Oriented Design and Programming Project, 30h, L1 (apprenticeship), IMT-Atlantique, Nantes, France

Master : Rémi Douence, Functional Programming with Haskell, 20h, M1 (apprenticeship), IMT-Atlantique, Nantes, France

Master : Rémi Douence, Introduction to scientific research in computer science (Project: an Haskell interpreter in Java), 45h, M2 (apprenticeship), IMT-Atlantique, Nantes, France

Licence : Hervé Grall, Algorithms and Discrete Mathematics, 25h , L3 (engineers), IMT-Atlantique, Nantes, France

Licence : Hervé Grall, Object Oriented Design and Programming, 25h , L3 (engineers), IMT-Atlantique, Nantes, France

Licence, Master : Hervé Grall, Modularity and Typing, 40h, L3 and M1, IMT-Atlantique, Nantes, France

Master : Hervé Grall, Service-oriented Computing, 40h, M1 and M2, IMT-Atlantique, Nantes, France

Master : Hervé Grall, Research Project - (Linear) Logic Programming in Coq, 90h (1/3 supervised), M1 and M2, IMT-Atlantique, Nantes, France

Licence : Guilhem Jaber, Computer Tools for Science, 36h, L1, Université de Nantes France

Master : Guilhem Jaber, Verification and Formal Proofs, 18h, M1, Université de Nantes, France

Master : Nicolas Tabareau, Homotopy Type Theory, 24h, M2 LMFI, Université Paris Diderot, France

8.2.2. Supervision

PhD : Gaetan Gilbert, A new foundation for the Coq proof assistant based on the insight of Homotopy Type Theory, IMT Atlantique, advisors: Matthieu Sozeau and Nicolas Tabareau

PhD : Ambroise Lafont,Towards an unbiased approach to specify, implement, and prove properties on programming languages, IMT Atlantique, advisors: Tom Hirschowitz and Nicolas Tabareau

PhD in progress: Xavier Montillet, Rewriting theory for effects and dependent types, Univ Nantes, advisors: Guillaume Munch-Maccagnoni and Nicolas Tabareau

PhD in progress: Théo Winterhalter, Extending the flexibility of the universe hierarchy in type theory, Univ Nantes, advisors: Matthieu Sozeau and Nicolas Tabareau

PhD in progress: Joachim Hotonnier, Deep Specification for Domain-Specific Modelling, advisors: Gerson Sunye (Naomod team), Massimo Tisi (Naomod team), Hervé Grall

PhD in progress: Igor Zhirkov, Certified Refactoring of C in the Coq proof assistant, advisors: Rémi Douence and Julien Cohen.

8.2.2.1. Supervision of interns

E. Bauer has visited the team for an L3 research internship from June to July on the subject “Categorical models of differential linear logic”, supervised by Marie Kerjean.

A. Ben Mansour has visited the team for an L3 research internship from June to August on the subject “Parametricity for languages with effects”, supervised by G. Munch-Maccagnoni.

M. Bertrand has visited the team from February to July for an internship on the subject “Effects in Type Theory”, supervised by N. Tabareau.

L. Pujet has visited the team from April to August for an internship on the subject “Interpreting Cubical Type Theory using forcing”, supervised by N. Tabareau.

G. Combette is visiting the team from October 2019 to February 2020 for an internship on the subject “Axiomatic denotational semantics for resource management in systems programming”, supervised by G. Munch-Maccagnoni.

L. Escot has visited the team from October to December for an internship on the subject “Univalent Parametricity at Scale”, supervised by N. Tabareau.

P. Geneau de Lamarlière has visited the team for an L3 research internship from June to July on the subject “Symbolic computations in algebraic number theory”, co-supervised by A. Mahboubi and S. Dahmen (VU Amsterdam).

8.2.3. Juries

- N. Tabareau has served as external member on the PhD jury of Kenji Maillard, defended November 25th at Inria Paris - Université Paris Sciences et Lettres.
- A. Mahboubi has served as external member on the PhD jury of Florian Faissole, defended December 13th at Paris Saclay University.
- A. Mahboubi has served as external member on the PhD jury of Gaëtan Gilbert, defended December 20th at IMT Atlantique.
- A. Mahboubi has served on the 2019 jury for recruiting Inria CRCN at Inria Rennes Bretagne Atlantique.
- A. Mahboubi has served on the 2019 jury for recruiting an “Agrégé Préparateur” at ENS Rennes.

8.3. Popularization

8.3.1. Education

- Hervé Grall has contributed to the project Merite, which aims to promote science learning in middle and high schools. He is the main contributor to the theme "Communication between machines". The project is coordinated by IMT-Atlantique in partnership with 7 other french higher education institutions, the rectorates of the Nantes and Rennes academies, and financed by the "Investments in the Future" and the fund FEDER Pays-de-la-Loire.
- A. Mahboubi has worked with composer Alessandro Bossetti and students of the professional high-school Lycée Michelet, on a project “Art and Mathematics”, supported by the Athenor theater.

8.3.2. Interventions

- P.-M. Pédrot was invited to give a talk about his scientific activities in Le Pleynet, during the "Semaine Sport-Études" of the first year students in computer science from the ENS Lyon.

- P.-M. Pédrot gave a similar talk about his scientific activities at the LS2N, during the visit of first year students in computer science from the ENS Cachan.

8.3.3. Internal action

- A. Mahboubi participates to the Irisa/Inria mentoring program.

9. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] B. AHRENS, A. HIRSCHOWITZ, A. LAFONT, M. MAGGESI.*Reduction Monads and Their Signatures*, in "Proceedings of the ACM on Programming Languages", January 2020 [*DOI :* 10.1145/3371099], https://hal.inria.fr/hal-02380682

[2] Z. CHENG, M. TISI, R. DOUENCE.*CoqTL: A Coq DSL for Rule-Based Model Transformation*, in "Software and Systems Modeling", 2019, p. 1-15, forthcoming [*DOI :* 10.1007/s10270-019-00765-6], https://hal.archives-ouvertes.fr/hal-02333564

[3] G. GILBERT, J. COCKX, M. SOZEAU, N. TABAREAU.*Definitional Proof-Irrelevance without K*, in "Proceedings of the ACM on Programming Languages", January 2019, p. 1-28 [*DOI :* 10.1145/329031610.1145/3290316], https://hal.inria.fr/hal-01859964

[4] G. JABER.*SyTeCi: Automating Contextual Equivalence for Higher-Order Programs with References*, in "Proceedings of the ACM on Programming Languages", 2019, vol. 28, p. 1-28, forthcoming [*DOI :* 10.1145/3371127], https://hal.archives-ouvertes.fr/hal-02388621

[5] A. MAHBOUBI, G. MELQUIOND, T. SIBUT-PINOTE.*Formally Verified Approximations of Definite Integrals*, in "Journal of Automated Reasoning", February 2019, vol. 62, n^o^ 2, p. 281-300 [*DOI :* 10.1007/s10817-018-9463-7], https://hal.inria.fr/hal-01630143

[6] P.-M. PÉDROT, N. TABAREAU, H. J. FEHRMANN, É. TANTER.*A Reasonably Exceptional Type Theory*, in "Proceedings of the ACM on Programming Languages", August 2019, vol. 3, p. 1-29 [*DOI :* 10.1145/3341712], https://hal.inria.fr/hal-02189128

[7] P.-M. PÉDROT, N. TABAREAU.*The Fire Triangle : How to Mix Substitution, Dependent Elimination, and Effects*, in "Proceedings of the ACM on Programming Languages", January 2020 [*DOI :* 10.1145/3371126], https://hal.archives-ouvertes.fr/hal-02383109

[8] M. SOZEAU, S. BOULIER, Y. FORSTER, N. TABAREAU, T. WINTERHALTER.*Coq Coq Correct! Verification of Type Checking and Erasure for Coq, in Coq*, in "Proceedings of the ACM on Programming Languages", January 2020 [*DOI :* 10.1145/3371076], https://hal.archives-ouvertes.fr/hal-02380196

[9] N. TABAREAU, É. TANTER.*Chemical foundations of distributed aspects*, in "Distributed Computing", June 2019, vol. 32, n^o^ Issue 3, p. 193–216, forthcoming [*DOI :* 10.1007/s00446-018-0334-6], https://hal.inria.fr/hal-01811884

International Conferences with Proceedings

[10] B. AHRENS, A. HIRSCHOWITZ, A. LAFONT, M. MAGGESI. *Modular specification of monads through higher-order presentations*, in "FSCD 2019 - 4th International Conference on Formal Structures for Computation and Deduction", Dortmund, Germany, June 2019, p. 1-16, https://arxiv.org/abs/1903.00922 - 17 pages [*DOI :* 10.4230/LIPIcs.FSCD.2019.6], https://hal.archives-ouvertes.fr/hal-02307998

[11] T. ALTENKIRCH, S. BOULIER, A. KAPOSI, N. TABAREAU. *Setoid type theory - a syntactic translation*, in "MPC 2019 - 13th International Conference on Mathematics of Program Construction", Porto, Portugal, LNCS, Springer, October 2019, vol. 11825, p. 155-196 [*DOI :* 10.1007/978-3-030-33636-3_7], https://hal.inria.fr/hal-02281225

[12] F. BRÉHARD, A. MAHBOUBI, D. POUS. *A certificate-based approach to formally verified approximations*, in "ITP 2019 - Tenth International Conference on Interactive Theorem Proving", Portland, United States, 2019, p. 1-19 [*DOI :* 10.4230/LIPIcs.ITP.2019.8], https://hal.laas.fr/hal-02088529

[13] É. MIQUEY. *Revisiting the duality of computation: an algebraic analysis of classical realizability models*, in "CSL 2020", Barcelone, Spain, LIPIcs, CSL 2020, January 2020, vol. 152, https://arxiv.org/abs/1910.02732 , https://hal.archives-ouvertes.fr/hal-02305560

[14] A. MÖRTBERG, L. PUJET. *Cubical Synthetic Homotopy Theory*, in "CPP 2020 - 9th ACM SIGPLAN International Conference on Certified Programs and Proofs", New Orleans, United States, ACM, January 2020 [*DOI :* 10.1145/3372885.3373825], https://hal.archives-ouvertes.fr/hal-02394145

[15] T. WINTERHALTER, M. SOZEAU, N. TABAREAU. *Eliminating Reflection from Type Theory : To the Legacy of Martin Hofmann*, in "CPP 2019 - 8th ACM SIGPLAN International Conference on Certified Programs and Proofs", Lisbonne, Portugal, ACM, January 2019, p. 91-103 [*DOI :* 10.1145/3293880.3294095], https://hal.archives-ouvertes.fr/hal-01849166

Conferences without Proceedings

[16] C. CHENAVIER, M. LUCAS. *The Diamond Lemma for non-terminating rewriting systems using deterministic reduction strategies*, in "IWC 2019 - 8th International Workshop on Confluence", Dortmund, Germany, June 2019, p. 1-5, https://hal.archives-ouvertes.fr/hal-02385139

[17] G. MUNCH-MACCAGNONI, R. DOUENCE. *Efficient Deconstruction with Typed Pointer Reversal (abstract)*, in "ML 2019 - Workshop", Berlin, Germany, KC Sivaramakrishnan, 2019, p. 1-8, https://hal.inria.fr/hal-02177326

Research Reports

[18] I. ZHIRKOV, J. COHEN, R. DOUENCE. *Memory bijections: reasoning about exact memory transformations induced by refactorings in CompCert C*, LS2N, Université de Nantes, March 2019, https://hal.archives-ouvertes.fr/hal-02078356

Other Publications

[19] D. ARA, M. LUCAS. *The folk model category structure on strict ω-categories is monoidal*, 2019, https://arxiv.org/abs/1909.13564 - 62 pages, https://hal.archives-ouvertes.fr/hal-02386617

[20] E. BAUER, M. KERJEAN.*Chiralités et exponentielles: un peu de différentiation*, December 2019, working paper or preprint, https://hal.inria.fr/hal-02320704

[21] H. HERBELIN, É. MIQUEY.*Continuation-and-environment-passing style translations: a focus on call-by-need*, January 2019, working paper or preprint, https://hal.inria.fr/hal-01972846

[22] A. HIRSCHOWITZ, T. HIRSCHOWITZ, A. LAFONT.*Modules over monads and operational semantics*, October 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02338144

[23] M. KERJEAN.*Chiralities in topological vector spaces*, December 2019, working paper or preprint, https://hal.inria.fr/hal-02334917

[24] M. KERJEAN, J.-S. LEMAY.*Higher-order distributions for differential linear logic*, January 2019, working paper or preprint, https://hal.inria.fr/hal-01969262

[25] M. LUCAS.*An implementation of polygraphs*, 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02385110

[26] M. SOZEAU, A. ANAND, S. BOULIER, C. COHEN, Y. FORSTER, F. KUNZE, G. MALECHA, N. TABAREAU, T. WINTERHALTER.*The MetaCoq Project*, June 2019, working paper or preprint, https://hal.inria.fr/hal-02167423

References in notes

[27] A. ABEL, T. COQUAND.*Untyped Algorithmic Equality for Martin-Löf's Logical Framework with Surjective Pairs*, in "Typed Lambda Calculi and Applications", P. URZYCZYN (editor), Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2005, vol. 3461, p. 23-38, http://dx.doi.org/10.1007/11417170_4

[28] B. ACCATTOLI, G. GUERRIERI.*Open Call-by-Value*, in "Programming Languages and Systems", January 2016, http://dx.doi.org/10.1007/978-3-319-47958-3_12

[29] D. AHMAN, N. GHANI, G. D. PLOTKIN.*Dependent Types and Fibred Computational Effects*, in "Proc. FoSSaCS", 2015

[30] A. AJOULI, J. COHEN, J.-C. ROYER.*Transformations between Composite and Visitor Implementations in Java*, in "Software Engineering and Advanced Applications (SEAA), 2013 39th EUROMICRO Conference on", Sept 2013, p. 25–32, http://dx.doi.org/10.1109/SEAA.2013.53

[31] J. ALDRICH.*The power of interoperability: why objects are inevitable*, in "ACM Symposium on New Ideas in Programming and Reflections on Software, Onward! 2013, part of SPLASH '13, Indianapolis, IN, USA, October 26-31, 2013", A. L. HOSKING, P. T. EUGSTER, R. HIRSCHFELD (editors), ACM, 2013, p. 101–116

[32] T. ALTENKIRCH, C. MCBRIDE, W. SWIERSTRA.*Observational equality, now!*, in "Proceedings of the ACM Workshop on Programming Languages meets Program Verification (PLPV 2007)", Freiburg, Germany, October 2007, p. 57–68

[33] J.-M. ANDREOLI.*Logic Programming with Focusing Proof in Linear Logic*, in "Journal of Logic and Computation", 1992, vol. 2, n° 3, p. 297-347

[34] E. ARAFAILOVA, N. BELDICEANU, R. DOUENCE, M. CARLSSON, P. FLENER, M. A. F. RODRÍGUEZ, J. PEARSON, H. SIMONIS.*Global Constraint Catalog, Volume II, Time-Series Constraints*, in "CoRR", 2016, vol. abs/1609.08925, http://arxiv.org/abs/1609.08925

[35] E. ARAFAILOVA, N. BELDICEANU, R. DOUENCE, P. FLENER, M. A. F. RODRÍGUEZ, J. PEARSON, H. SIMONIS.*Time-Series Constraints: Improvements and Application in CP and MIP Contexts*, in "Integration of AI and OR Techniques in Constraint Programming - 13th International Conference, CPAIOR 2016, Banff, AB, Canada, May 29 - June 1, 2016, Proceedings", C. QUIMPER (editor), Lecture Notes in Computer Science, Springer, 2016, vol. 9676, p. 18–34, https://doi.org/10.1007/978-3-319-33954-2_2

[36] A. F. BARCO, J. FAGES, É. VAREILLES, M. ALDANONDO, P. GABORIT.*Open Packing for Facade-Layout Synthesis Under a General Purpose Solver*, in "Principles and Practice of Constraint Programming - 21st International Conference, CP 2015, Cork, Ireland, August 31 - September 4, 2015, Proceedings", G. PESANT (editor), Lecture Notes in Computer Science, Springer, 2015, vol. 9255, p. 508–523, http://dx.doi.org/10.1007/978-3-319-23219-5_36

[37] N. BELDICEANU, M. CARLSSON, R. DOUENCE, H. SIMONIS.*Using finite transducers for describing and synthesising structural time-series constraints*, in "Constraints", 2016, vol. 21, n^{o} 1, p. 22–40, http://dx.doi.org/10.1007/s10601-015-9200-3

[38] S. BERARDI, M. BEZEM, T. COQUAND.*On the computational content of the axiom of choice*, in "The Journal of Symbolic Logic", 1998, vol. 63, n^{o} 02, p. 600–622

[39] M. BEZEM, T. COQUAND, S. HUBER.*A model of type theory in cubical sets*, in "Preprint, September", 2013

[40] S. BOULIER, P.-M. PÉDROT, N. TABAREAU.*The next 700 syntactical models of type theory*, in "Certified Programs and Proofs (CPP 2017)", Paris, France, January 2017, p. 182 - 194 [*DOI :* 10.1145/3018610.3018620], https://hal.inria.fr/hal-01445835

[41] E. BRADY.*Idris, a general-purpose dependently typed programming language: Design and implementation*, in "J. Funct. Program.", 2013, vol. 23, n^{o} 5, p. 552–593, https://doi.org/10.1017/S095679681300018X

[42] F. CHYZAK, A. MAHBOUBI, T. SIBUT-PINOTE, E. TASSI.*A Computer-Algebra-Based Formal Proof of the Irrationality of ζ(3)*, in "Interactive Theorem Proving", R. G. GERWIN KLEIN (editor), Lecture Notes in Computer Science, Springer, 2014, vol. 8558

[43] J. COHEN, A. AJOULI.*Practical Use of Static Composition of Refactoring Operations*, in "Proceedings of the 28th Annual ACM Symposium on Applied Computing", SAC '13, ACM, 2013, p. 1700–1705, http://dx.doi.org/10.1145/2480362.2480684

[44] J. COHEN.*Renaming Global Variables in C Mechanically Proved Correct*, in "Proceedings of the Fourth International Workshop on Verification and Program Transformation, Eindhoven, The Netherlands, 2nd April 2016", G. HAMILTON, A. LISITSA, A. P. NEMYTYKH (editors), Electronic Proceedings in Theoretical Computer Science, Open Publishing Association, 2016, vol. 216, p. 50-64, http://dx.doi.org/10.4204/EPTCS.216.3

[45] C. COHEN, T. COQUAND, S. HUBER, A. MÖRTBERG.*Cubical Type Theory: a constructive interpretation of the univalence axiom*, 2016, To appear in post-proceedings of Types for Proofs and Programs (TYPES 2015)

[46] P. COHEN, M. DAVIS.*Set theory and the continuum hypothesis*, WA Benjamin New York, 1966

[47] J. COHEN, R. DOUENCE, A. AJOULI.*Invertible Program Restructurings for Continuing Modular Maintenance*, in "Software Maintenance and Reengineering (CSMR), 2012 16th European Conference on", March 2012, p. 347-352, http://dx.doi.org/10.1109/CSMR.2012.42

[48] W. R. COOK.*On understanding data abstraction, revisited*, in "Proceedings of the 24th Annual ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications, OOPSLA 2009, October 25-29, 2009, Orlando, Florida, USA", S. ARORA, G. T. LEAVENS (editors), ACM, 2009, p. 557–572, http://doi.acm.org/10.1145/1640089.1640133

[49] COQ DEVELOPMENT TEAM, THE.*The Coq proof assistant reference manual*, 2015, Version 8.5, http://coq.inria.fr

[50] P.-L. CURIEN, M. FIORE, G. MUNCH-MACCAGNONI.*A Theory of Effects and Resources: Adjunction Models and Polarised Calculi*, in "Proc. POPL", 2016, http://dx.doi.org/10.1145/2837614.2837652

[51] P.-L. CURIEN, H. HERBELIN.*The duality of computation*, in "ACM SIGPLAN Notices", 2000, vol. 35, p. 233–243

[52] D. DELAHAYE, M. MAYERO.*Dealing with algebraic expressions over a field in Coq using Maple*, in "J. Symbolic Comput.", 2005, vol. 39, n^o 5, p. 569–592, Special issue on the integration of automated reasoning and computer algebra systems, http://dx.doi.org/10.1016/j.jsc.2004.12.004

[53] R. DOUENCE, X. LORCA, N. LORIANT.*Lazy Composition of Representations in Java*, in "Software Composition, 8th International Conference, SC 2009, Zurich, Switzerland, July 2-3, 2009. Proceedings", A. BERGEL, J. FABRY (editors), Lecture Notes in Computer Science, Springer, 2009, vol. 5634, p. 55–71, https://doi.org/10.1007/978-3-642-02655-3_6

[54] T. EHRHARD.*Call-by-push-value from a linear logic point of view*, in "European Symposium on Programming Languages and Systems", Springer, 2016, p. 202–228

[55] A. FRISCH, G. CASTAGNA, V. BENZAKEN.*Semantic Subtyping: Dealing Set-theoretically with Function, Union, Intersection, and Negation Types*, in "J. ACM", September 2008, vol. 55, n^o 4, p. 19:1–19:64

[56] C. FÜHRMANN.*Direct Models for the Computational Lambda Calculus*, in "Electr. Notes Theor. Comput. Sci.", 1999, vol. 20, p. 245-292

[57] J.-Y. GIRARD.*Linear Logic*, in "Theoretical Computer Science", 1987, vol. 50, p. 1-102

[58] J.-Y. GIRARD, A. SCEDROV, P. J. SCOTT.*Normal Forms and Cut-Free Proofs as Natural Transformations*, in "in : Logic From Computer Science, Mathematical Science Research Institute Publications 21", Springer-Verlag, 1992, p. 217–241

[59] G. GONTHIER, A. ASPERTI, J. AVIGAD, Y. BERTOT, C. COHEN, F. GARILLOT, S. ROUX, A. MAHBOUBI, R. O'CONNOR, S. OULD BIHA, I. PASCA, L. RIDEAU, A. SOLOVYEV, E. TASSI, L. THÉRY.*A Machine-Checked Proof of the Odd Order Theorem*, in "Interactive Theorem Proving", S. BLAZY, C. PAULIN-

MOHRING, D. PICHARDIE (editors), Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2013, vol. 7998, p. 163-179, http://dx.doi.org/10.1007/978-3-642-39634-2_14

[60] G. GONTHIER.*Formal proofs—the four-colour theorem*, in "Notices of the AMS", 2008, vol. 55, n° 11, p. 1382-1393

[61] G. GONTHIER, A. MAHBOUBI, E. TASSI.*A Small Scale Reflection Extension for the Coq system*, Inria, 2008, n° RR-6455, The Reference Manual of the Ssreflect extension to the Coq tactic language, available at http://hal.inria.fr/inria-00258384

[62] T. G. GRIFFIN.*A Formulae-as-Types Notion of Control*, in "Seventeenth Annual ACM Symposium on Principles of Programming Languages", ACM Press, 1990, p. 47–58

[63] Y. GUÉHÉNEUC, R. DOUENCE, N. JUSSIEN.*No Java without Caffeine: A Tool for Dynamic Analysis of Java Programs*, in "17th IEEE International Conference on Automated Software Engineering (ASE 2002), 23-27 September 2002, Edinburgh, Scotland, UK", IEEE Computer Society, 2002, 117, https://doi.org/10.1109/ASE.2002.1115000

[64] T. C. HALES, M. ADAMS, G. BAUER, D. T. DANG, J. HARRISON, T. L. HOANG, C. KALISZYK, V. MAGRON, S. MCLAUGHLIN, T. T. NGUYEN, T. Q. NGUYEN, T. NIPKOW, S. OBUA, J. PLESO, J. RUTE, A. SOLOVYEV, A. H. T. TA, T. N. TRAN, D. T. TRIEU, J. URBAN, K. K. VU, R. ZUMKELLER.*A formal proof of the Kepler conjecture*, in "CoRR", 2015, vol. abs/1501.02155, http://arxiv.org/abs/1501.02155

[65] H. A. HELFGOTT.*The ternary Goldbach conjecture is true*, in "ArXiv e-prints", December 2013

[66] H. HERBELIN.*A Constructive Proof of Dependent Choice, Compatible with Classical Logic*, in "LICS 2012 - 27th Annual ACM/IEEE Symposium on Logic in Computer Science", Dubrovnik, Croatia, IEEE Computer Society, June 2012, p. 365-374, https://hal.inria.fr/hal-00697240

[67] H. HERBELIN, É. MIQUEY.*Toward dependent choice: a classical sequent calculus with dependent types*, in "TYPES 2015", 2015

[68] T. HIRSCHOWITZ.*Cartesian closed 2-categories and permutation equivalence in higher-order rewriting*, in "Logical Methods in Computer Science", 2013, vol. 9, n° 3, 10, 19 pages [*DOI :* 10.2168/LMCS-9(3:10)2013], https://hal.archives-ouvertes.fr/hal-00540205

[69] F. IMMLER.*Verified Reachability Analysis of Continuous Systems*, in "Tools and Algorithms for the Construction and Analysis of Systems - 21st International Conference, TACAS 2015, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2015, London, UK, April 11-18, 2015. Proceedings", C. BAIER, C. TINELLI (editors), Lecture Notes in Computer Science, Springer, 2015, vol. 9035, p. 37–51

[70] G. JABER, G. LEWERTOWSKI, P.-M. PÉDROT, M. SOZEAU, N. TABAREAU.*The Definitional Side of the Forcing*, in "Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science, LICS '16, New York, NY, USA, July 5-8, 2016", 2016, p. 367–376

[71] G. JABER, N. TABAREAU, M. SOZEAU.*Extending type theory with forcing*, in "Logic in Computer Science (LICS), 2012", IEEE, 2012, p. 395–404

[72] C. B. JAY, N. GHANI.*The Virtues of Eta-Expansion*, in "J. Funct. Program.", 1995, vol. 5, n^{o} 2, p. 135-154

[73] U. KOHLENBACH.*Applied proof theory: proof interpretations and their use in mathematics*, Springer Science & Business Media, 2008

[74] J.-L. KRIVINE.*Realizability algebras II : new models of ZF + DC*, in "Logical Methods in Computer Science", 2012, vol. 8, n^{o} 1

[75] J. LAMBEK, P. J. SCOTT.*Introduction to higher order categorical logic*, Cambridge University Press, New York, NY, USA, 1986

[76] S. M. LANE, I. MOERDIJK.*Sheaves in Geometry and Logic*, Springer-Verlag, 1992

[77] R. LEPIGRE.*A classical realizability model for a semantical value restriction*, in "European Symposium on Programming Languages and Systems", Springer, 2016, p. 476–502

[78] X. LEROY.*Formal certification of a compiler back-end or: programming a compiler with a proof assistant*, in "ACM SIGPLAN Notices", 2006, vol. 41, n^{o} 1, p. 42–54

[79] P. B. LEVY.*Call-By-Push-Value: A Functional/Imperative Synthesis*, Semantic Structures in Computation, Springer, 2004, vol. 2

[80] P. B. LEVY.*Contextual isomorphisms*, in "Proceedings of the 44th ACM SIGPLAN Symposium on Principles of Programming Languages", ACM, 2017, p. 400–414

[81] C. LIANG, D. MILLER.*Focusing and polarization in linear, intuitionistic, and classical logics*, in "Theor. Comput. Sci.", 2009, vol. 410, n^{o} 46, p. 4747-4768

[82] Z. LUO, S. SOLOVIEV, T. XUE.*Coercive subtyping: Theory and implementation*, in "Inf. Comput.", 2013, vol. 223, p. 18–42

[83] J. LURIE.*Higher topos theory*, Annals of mathematics studies, Princeton University Press, Princeton, N.J., Oxford, 2009

[84] S. MAC LANE.*Natural associativity and commutativity*, in "Selected Papers", 1979, p. 415–433

[85] P. MARTIN-LÖF.*An intuitionistic theory of types: predicative part*, in "Logic Colloquium '73", 1975, vol. Studies in Logic and the Foundations of Mathematics, n^{o} 80, p. 73–118

[86] P.-A. MELLIÈS.*Asynchronous Games 3 An Innocent Model of Linear Logic*, in "Electr. Notes Theor. Comput. Sci.", 2005, vol. 122, p. 171-192

[87] P.-A. MELLIÈS.*Asynchronous Games 4: A Fully Complete Model of Propositional Linear Logic*, in "LICS", 2005, p. 386-395

[88] P.-A. MELLIÈS, N. TABAREAU.*Resource modalities in tensor logic*, in "Ann. Pure Appl. Logic", 2010, vol. 161, n^{o} 5, p. 632-653

[89] É. MIQUEY.*A classical sequent calculus with dependent types*, in "European Symposium on Programming", Springer, 2017, p. 777–803

[90] E. MOGGI.*Computational lambda-calculus and monads*, in "Proceedings of the Fourth Annual IEEE Symposium on Logic in Computer Science (LICS 1989)", IEEE Computer Society Press, June 1989, p. 14–23

[91] E. MOGGI.*Notions of computation and monads*, in "Inf. Comput.", July 1991, vol. 93, n^{o} 1, p. 55–92, http://dx.doi.org/10.1016/0890-5401(91)90052-4

[92] G. MUNCH-MACCAGNONI.*Formulae-as-Types for an Involutive Negation*, in "Proceedings of the joint meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (CSL-LICS)", 2014

[93] G. MUNCH-MACCAGNONI.*Models of a Non-Associative Composition*, in "Proc. FoSSaCS", A. MUSCHOLL (editor), LNCS, Springer, 2014, vol. 8412, p. 397–412

[94] G. MUNCH-MACCAGNONI.*Note on Curry's style for Linear Call-by-Push-Value*, 2017, https://hal.inria.fr/hal-01528857

[95] A. NANEVSKI, G. MORRISETT, A. SHINNAR, P. GOVEREAU, L. BIRKEDAL.*Ynot: Reasoning with the awkward squad*, 2008

[96] C. OKASAKI.*Purely functional data structures*, Cambridge University Press, 1999

[97] C. PRUD'HOMME, X. LORCA, R. DOUENCE, N. JUSSIEN.*Propagation engine prototyping with a domain specific language*, in "Constraints", 2014, vol. 19, n^{o} 1, p. 57–76, https://doi.org/10.1007/s10601-013-9151-5

[98] C. PRUD'HOMME.*Contrôle de la propagation et de la recherche dans un solveur de contraintes. (Controlling propagation and search within a constraint solver)*, École des mines de Nantes, France, 2014, https://tel.archives-ouvertes.fr/tel-01060921

[99] P.-M. PÉDROT, N. TABAREAU.*An Effectful Way to Eliminate Addiction to Dependence*, January 2017, https://hal.inria.fr/hal-01441829

[100] K. QUIRIN, N. TABAREAU.*Lawvere-Tierney sheafification in Homotopy Type Theory*, in "Journal of Formalized Reasoning", 2016, vol. 9, n^{o} 2 [*DOI :* 10.6092/ISSN.1972-5787/6232], https://hal.inria.fr/hal-01451710

[101] J. C. REYNOLDS.*Types, Abstraction and Parametric Polymorphism*, in "IFIP Congress", 1983, p. 513-523

[102] P. SELINGER.*Control Categories and Duality: On the Categorical Semantics of the Lambda-Mu Calculus*, in "Math. Struct in Comp. Sci.", 2001, vol. 11, n^{o} 2, p. 207–260

[103] S. G. SIMPSON.*Subsystems of Second Order Arithmetic*, Second, Cambridge University Press, 2009, Cambridge Books Online, http://dx.doi.org/10.1017/CBO9780511581007

[104] K. STØVRING.*Extending the Extensional Lambda Calculus with Surjective Pairing is Conservative*, in "Logical Methods in Computer Science", 2006, vol. 2, n^{o} 2

[105] N. SWAMY, C. HRIŢCU, C. KELLER, A. RASTOGI, A. DELIGNAT-LAVAUD, S. FOREST, K. BHARGAVAN, C. FOURNET, P.-Y. STRUB, M. KOHLWEISS, J.-K. ZINZINDOHOUE, S. ZANELLA-BÉGUELIN.*Dependent Types and Multi-Monadic Effects in F**, in "43nd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL)", ACM, January 2016, p. 256-270, https://www.fstar-lang.org/papers/mumon/

[106] É. TANTER, N. TABAREAU.*Gradual Certified Programming in Coq*, in "Proceedings of the 11th ACM Dynamic Languages Symposium (DLS 2015)", Pittsburgh, PA, USA, ACM Press, October 2015, p. 26–40

[107] UNIVALENT FOUNDATIONS PROJECT.*Homotopy Type Theory: Univalent Foundations for Mathematics*, http://homotopytypetheory.org/book, 2013

[108] M. VÁKÁR.*A Framework for Dependent Types and Effects*, in "arXiv preprint arXiv:1512.08009", 2015

[109] B. ZILIANI, D. DREYER, N. R. KRISHNASWAMI, A. NANEVSKI, V. VAFEIADIS.*Mtac: A monad for typed tactic programming in Coq*, in "Journal of Functional Programming", 2015, vol. 25, http://dx.doi.org/10.1017/S0956796815000118

[110] F. VAN RAAMSDONK.*Higher-order Rewriting*, in "Proc. Rewrit. Tech. App.", LNCS, Springer, 1999, vol. 1631, p. 220-239

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

Project-Team GENSCALE

Creation of the Team: 2012 January 01, updated into Project-Team: 2013 January 01

Keywords:

Computer Science and Digital Science:

A1.1.1. - Multicore, Manycore
A1.1.2. - Hardware accelerators (GPGPU, FPGA, etc.)
A1.1.3. - Memory models
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:

B1.1.4. - Genetics and genomics
B1.1.7. - Bioinformatics
B2.2.6. - Neurodegenerative diseases
B3.6. - Ecology
B3.6.1. - Biodiversity

1. Team, Visitors, External Collaborators

Research Scientists

Dominique Lavenier [Team leader, CNRS, Senior Researcher, HDR]
Claire Lemaitre [Inria, Researcher]
Jacques Nicolas [Inria, Senior Researcher, HDR]
Pierre Peterlongo [Inria, Researcher, HDR]

Faculty Members

Roumen Andonov [Univ de Rennes I, Professor, HDR]
Emeline Roux [Univ de Lorraine, Associate Professor]

External Collaborators

Susete Alves Carvalho [INRA]
Fabrice Legeai [INRA]
Mohammed Amin Madoui [CEA, from Sep 2019]

Technical Staff

Charles Deltel [Inria, Engineer]
Anne Guichard [INRA, Engineer, from Jul 2019]
Gwendal Virlet [CNRS, Engineer, until Sep 2019]

PhD Students

Kévin Da Silva [Inria, PhD Student]
Wesley Delage [Inria, PhD Student]
Clara Delahaye [Univ de Rennes I, PhD Student, from Oct 2019]
Sébastien Francois [Univ de Rennes I, PhD Student, until Sep 2019]

Rati Kar [INRA, PhD Student, until Jan 2019]
Lolita Lecompte [Inria, PhD Student]
Téo Lemane [Inria, PhD Student, from Oct 2019]
Grégoire Siekaniec [INRA, PhD Student]

Post-Doctoral Fellow
Céline Le Beguec [Inria, Post-Doctoral Fellow]

Administrative Assistants
Marie Le Roic [Inria, Administrative Assistant, from Dec 2019]
Marie Le Roic [Univ de Rennes I, Administrative Assistant, until Nov 2019]

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 sequencing technologies, especially next generation sequencing (NGS), that provide billions of very short DNA fragments of high quality, and third generation sequencing (TGS), that provide millions of long DNA fragments of lower quality. NGS and TGS techniques bring 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 throughput 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 sequenced 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 sequencing 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 tools dedicated to sequencing data. 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 or ITN European 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 Structures

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, or multiple genomes coming from a same sample as in metagenomics, require high computing resources, and more specifically very important memory configuration. The last advances in TGS technologies bring also new challenges to represent or search information based on sequencing data with high error rate.

Part of 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. 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 [3], [4].

A correlated research direction 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 [7].

3.2. Axis 2: Algorithms

The main goal of the GenScale team is to develop optimized tools dedicated to genomic data 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 genomic data processing 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 [1].
- **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 [3] and on the scaffolding step [5]. Both NGS and TGS technologies are taken into consideration, either independently or using combined approaches.
- **Detection of variants** This is often the main information one wants to extract from the sequencing data. Variants range from SNPs or short indels to structural variants that are large insertions/deletions and long inversions over the chromosomes. We developed original methods to find variants without any reference genome [10], to detect structural variants using local NGS assembly approaches [9] or TGS processing.

- **Metagenomics** We focused 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 [2].

3.3. Axis 3: Parallelism

This third axis investigates a supplementary way to increase performances and scalability of genomic 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 [4]. The second is a parallel algorithmic approach: during the development step, the algorithms are constrained by parallel criteria [2]. 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. Bioinformatics analysis can be based on the detection of SNPs (Single Nucleotide Polymorphism) from a set of predefined target genes. One can also scan the complete genome and report all kinds of mutations, including complex mutations such as large insertions or deletions, that could be associated with genetic or cancer diseases.

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 genetic 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

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 [6].

Improving plant breeding: Such projects aim at identifying favorable alleles at loci contributing to phenotypic variation, characterizing polymorphism at the functional level and 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.

4.4. Environment

Food quality control: One way to check food contaminated with bacteria is to extract DNA from a product and identify the different strains it contains. This can now be done quickly with low-cost sequencing technologies such as the MinION sequencer from Oxford Nanopore Technologies.

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 its role, for example, in the CO2 sequestration.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

The Gilles Kahn *accessits* **prize was awarded to Camille Marchet** for her PhD thesis: *From reads to transcripts: de novo methods for the analysis of transcriptome second and third generation sequencing* [8]. This thesis was prepared in the GenScale team under the supervision of P. Peterlongo.

The Gilles Kahn prize is awarded each year by the SIF, the French Society of Computer Science, for an excellent PhD thesis in the field of computer science.

The thesis of Camille dealt with the processing of transcriptome sequencing data. More precisely, the question was how to take advantage of the characteristics of the data produced by third generation sequencing technologies, as they produce large sequences covering the total length of RNA molecules. The core work of this thesis consisted in the methodological development and implementation of new algorithms allowing the clustering of third generation sequences by gene, then their correction and finally the detection of the different isoforms of each gene.

6. New Software and Platforms

6.1. SVJedi

KEYWORDS: High throughput sequencing - Structural Variation - Genome analysis

FUNCTIONAL DESCRIPTION: SVJedi is a structural variation (SV) genotyper for long read data. Based on a representation of the different alleles, it estimates the genotype of each variant in a given individual sample based on allele-specific alignment counts. SVJedi takes as input a variant file (VCF), a reference genome (fasta) and a long read file (fasta/fastq) and outputs the initial variant file with an additional column containing genotyping information (VCF).

- Participants: Claire Lemaitre, Lolita Lecompte, Pierre Peterlongo and Dominique Lavenier
- Contact: Claire Lemaitre
- URL: https://github.com/llecompte/SVJedi

6.2. MinYS

MineYourSymbiont

KEYWORDS: High throughput sequencing - Genome assembly - Metagenomics

FUNCTIONAL DESCRIPTION: MinYS allows targeted assembly of a bacterial genome of interest in a metagenomic short read sequencing sample using a reference-guided pipeline. First, taking advantage of a potentially distant reference genome, a subset of the metagenomic reads is assembled into a set of backbone contigs. Then, this first draft assembly is completed using the whole metagenomic readset in a de novo manner. The resulting assembly is output as a genome graph, allowing to distinguish different strains with potential structural variants coexisting in the sample.

- Contact: Claire Lemaitre
- URL: https://github.com/cguyomar/MinYS

6.3. Simka

KEYWORDS: Comparative metagenomics - K-mer - Distance - Ecology

FUNCTIONAL DESCRIPTION: Simka is a comparative metagenomics method dedicated to NGS datasets. It computes a large collection of distances classically used in ecology to compare communities by approximating species counts by k-mer counts. The method scales to a large number of datasets thanks to an efficient and parallel kmer-counting strategy that processes all datasets simultaneoulsy. SimkaMin is distributed also with Simka. SimkaMin is a faster and more resource-frugal version of Simka. It outputs approximate (but very similar) results by subsampling the kmer space.

RELEASE FUNCTIONAL DESCRIPTION: Since release version 1.5.0, SimkaMin is distributed alongside Simka. SimkaMin is also a de novo comparative metagenomics tool. It is a faster and more resource-frugal version of Simka. It outputs approximate (but very similar) results as Simka by subsampling the kmer space. With this strategy, and with default parameters, SimkaMin is an order of magnitude faster, uses 10 times less memory and 70 times less disk than Simka.

- Participants: Claire Lemaitre, Dominique Lavenier, Gaëtan Benoit and Pierre Peterlongo
- Contact: Claire Lemaitre
- URL: https://gatb.inria.fr/software/simka/

6.4. 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.

RELEASE FUNCTIONAL DESCRIPTION: * Substantive improvements: better quality of results (accuracy and recall), better filtering of obtained results * Formal improvements: better organization of scripts, better presentation of results

- Participants: Pierre Peterlongo and Claire Lemaitre
- Contact: Pierre Peterlongo
- URL: https://github.com/GATB/DiscoSnp

7. New Results

7.1. Algorithms & Methods

7.1.1. SV genotyping

Participants: Dominique Lavenier, Lolita Lecompte, Claire Lemaitre, Pierre Peterlongo.

Structural variations (SV) are genomic variants of at least 50 base pairs (bp) that can be rearranged within the genome and thus can have a major impact on biological processes. Sequencing data from third generation technologies have made it possible to better characterize SVs. Although many SV callers have been published recently, there is no published method to date dedicated to genotyping SVs with this type of data. Variant genotyping consists in estimating the presence and ploidy or absence of a set of known variants in a newly sequenced individual. Thus, in this paper, we present a new method and its implementation, SVJedi, to genotype SVs with long reads. From a set of known SVs and a reference genome, our approach first generates local sequences representing the two possible alleles for each SV. Long read data are then aligned to these generated sequences and a careful analysis of the alignments consists in identifying only the informative ones to estimate the genotype for each SV. SVJedi achieves high accuracy on simulated and real human data and we demonstrate its substancial benefits with respect to other existing approaches, namely SV discovery with long reads and SV genotyping with short reads [23], [24], [35]. SVJedi is implemented in Python and available at https://github.com/llecompte/SVJedi.

7.1.2. Genome assembly of targeted organisms in metagenomic data

Participants: Wesley Delage, Fabrice Legeai, Claire Lemaitre.

In this work, we propose a two-step targeted assembly method tailored for metagenomic data, called MinYS (for MineYourSymbiont). First, a subset of the reads belonging to the species of interest are recruited by mapping and assembled *de novo* into backbone contigs using a classical assembler. Then an all-versus-all contig gap-filling is performed using a novel version of MindTheGap with the whole metagenomic dataset. The originality and success of the approach lie in this second step, that enables to assemble the missing regions between the backbone contigs, which may be regions absent or too divergent from the reference genome. The result of the method is a genome assembly graph in gfa format, accounting for the potential structural variations identified within the sample. We showed that MinYS is able to assemble the *Buchnera aphidicola* genome in a single contig in pea aphid metagenomic samples, even when using a divergent reference genome, it runs at least 10 times faster than classical *de novo* metagenomics assemblers and it is able to recover large structural variations co-existing in a sample. MinYS is a Python3 pipeline, distributed on github (https://github.com/cguyomar/MinYS) and as a conda package in the bio-conda repository [22].

7.1.3. SimkaMin: subsampling the kmer space for efficient comparative metagenomics

Participants: Claire Lemaitre, Pierre Peterlongo.

SimkaMin [12] is a quick comparative metagenomics tool with low disk and memory footprints, thanks to an efficient data subsampling scheme used to estimate Bray-Curtis and Jaccard dissimilarities. One billion metagenomic reads can be analyzed in less than 3 minutes, with tiny memory (1.09 GB) and disk ($\sim$0.3 GB) requirements and without altering the quality of the downstream comparative analyses, making of SimkaMin a tool perfectly tailored for very large-scale metagenomic projects.

7.1.4. Haplotype reconstruction: phasing co-localized variants

Participants: Mohammed Amin Madoui, Pierre Peterlongo.

In collaboration with Amin Madoui from the Genoscope (CEA), we develop a new methodology to reconstruct haplotypes or strain genomes directly from raw sequencing set of (metagenomic) reads. The goal is to propose long assembled sequences (i.e. complete genomes are not mandatory) such that each assembled sequence belongs to only one sequenced chromosome and is not a consensus of several similar sequences. Downstream, this enables to perform population genomics analyses.

The key idea is to use the DiscoSnp [10] output, detecting set of variant alleles that are co-localized on input reads or pairs of input reads. Then we finally reconstruct set of sequences that are as parsimonious as possible with those observations.

7.1.5. Finding all maximal perfect haplotype blocks in linear time

Participant: Pierre Peterlongo.

Recent large-scale community sequencing efforts allow at an unprecedented level of detail the identification of genomic regions that show signatures of natural selection. However, traditional methods for identifying such regions from individuals' haplotype data require excessive computing times and therefore are not applicable to current datasets. In 2019, Cunha et al. (Proceedings of BSB 2019) suggested the maximal perfect haplotype block as a very simple combinatorial pattern, forming the basis of a new method to perform rapid genome-wide selection scans. The algorithm they presented for identifying these blocks, however, had a worst-case running time quadratic in the genome length. It was posed as an open problem whether an optimal, linear-time algorithm exists. We gave two algorithms that achieve this time bound, one which is conceptually very simple and uses suffix trees and a second one using the positional Burrows-Wheeler Transform, that is very efficient also in practice [20].

7.1.6. Short read correction

Participant: Pierre Peterlongo.

We propose a new approach for the correction of NGS reads. This approach is based on the construction of a clean de Bruijn graph in which the correction is made at the contig level. In a second step, original reads are mapped on this graph, allowing to correct the original reads [16].

7.1.7. Large-scale kmer indexation

Participants: Téo Lemane, Pierre Peterlongo.

In the SeqDigger ANR project framework (see dedicated Section), we aim to index TB or PB of genomic sequences, assembled or not. The central idea is to assign any kmer (word of length k) to the set of indexed dataset it belongs to. For doing this we have proposed a method that improves one of the state of the art algorithm (HowDeSBT [38]) by optimizing the way kmers are counted and represented [36].

7.1.8. Proteogenomics workflow for the expert annotation of eukaryotic genomes

Participant: Pierre Peterlongo.

Accurate structural annotation of genomes is still a challenge, despite the progress made over the past decade. The prediction of gene structure remains difficult, especially for eukaryotic species, and is often erroneous and incomplete. In [15], we proposed a proteogenomics strategy, taking advantage of the combination of proteomics datasets and bioinformatics tools, to identify novel protein coding-genes and splice isoforms, to assign correct start sites, and to validate predicted exons and genes.

7.1.9. Gap-filling with linked-reads data

Participants: Anne Guichard, Fabrice Legeai, Claire Lemaitre, Arthur Le Bars, Pierre Peterlongo.

We develop a novel approach for filling assembly gaps with linked reads data (typically 10X Genomics technology). The approach is based on local assembly using our tool MindTheGap [9], and takes advantage of barcode information to reduce the input read set in order to reduce the de Bruijn graph complexity. The approach is applied to recover the genomic structure of a 1.3 Mb locus of interest in a dozen of re-sequenced butterfly genomes (*H. numata*) in the Supergene ANR project context.

7.1.10. Statistically Significant Discriminative Patterns Searching

Participants: Gwendal Virlet, Dominique Lavenier.

We propose a novel algorithm, called SSDPS, to discover patterns in two-class datasets. The algorithm, developed in collaboration with the LACODAM Inria team, owes its efficiency to an original enumeration strategy of the patterns, which allows to exploit some degrees of anti-monotonicity on the measures of discriminance and statistical significance. Experimental results demonstrate that the performance of the algorithm is better than others. In addition, the number of generated patterns is much less than the number of the other algorithms. An experiment on real data also shows that SSDPS efficiently detects multiple SNPs combinations in genetic data [27].

7.2. Optimisation

7.2.1. Chloroplast genome assembly

Participants: Sébastien Francois, Roumen Andonov, Dominique Lavenier.

This research focuses on the last two stages of *de novo* genome assembly, namely, scaffolding and gap-filling, and shows that they can be solved as part of a single optimization problem. Our approach is based on modeling genome assembly as a problem of finding a simple path in a specific graph that satisfies as many distance constraints as possible encoding the read-pair insert-size information. We formulate it as a mixed-integer linear programming (MILP) problem and apply an optimization solver to find the exact solutions on a benchmark of chloroplast genomes. We show that the presence of repetitions in the set of unitigs is the main reason for the existence of multiple equivalent solutions that are associated to alternative subpaths. We also describe two sufficient conditions and we design efficient algorithms for identifying these subpaths. Comparisons of the results achieved by our tool with the ones obtained with recent assemblers are also presented [11].

7.2.2. Integer Linear Programming for Metabolic Networks

Participants: Kerian Thuillier, Roumen Andonov.

Metabolic networks are a helpful tool to represent and study cell metabolisms. They contain information about every reaction occurring inside an organism. However, metabolic networks of poorly studied species are often incomplete. It is possible to complete these networks with knowledge of other well-known species.

In this study, we present a new linear programming approach for the problem of topological activation in metabolic networks based on flows and the Miller, Tucker and Zemlin (MTZ) formulation for solving the longest path problem. We developed a tool called *Flutampl* with AMPL (A Mathematical Programming Language). It returns optimal solutions for the hybrid completion directly from *sbml* files (the data format used for modelling metabolic networks) [37].

7.2.3. Integer Linear Programming for De novo Long Reads Assembly

Participants: Victor Epain, Roumen Andonov, Dominique Lavenier.

To tackle the de novo long read assembly problem, we investigate a new 2-step method based on integer linear programming. The first step orders the long reads and the second one generates a consensus sequence. Each step is based on a different IPL specification. In 2019, we focused on step 1: long reads are first compared to build an overlapping graph. Then we use integer linear programming to find the heaviest path in a graph $G = (V, E, \lambda)$, where V is the vertices set corresponding to the long reads, E the edge set associated to the overlaps between long reads and λ the overlap length. For large graph, V is partitioned into several parts, each one is solved independently, and the solutions are merged together. Preliminary experimentation show that bacteria assemblies can be successfully solved in a few minutes [31].

7.3. Experiments with the MinION Nanopore sequencer

7.3.1. Storing information on DNA Molecules

Participants: Dominique Lavenier, Emeline Roux, Ariane Badoual.

In 2019, we started a new research activity aiming to explore the possibility to use the DNA molecules as a storage medium. We designed a complete DNA storage system based on the Oxford Nanopore sequencing technology and performed several experimentations by sequencing several synthetized DNA fragments ranging from 500 to 1,000 bp. These sequences have been designed with ad-hoc coding to prevent specific sequencing errors of the nanopore technology such as indel errors in homo-polymer sequences [29] [34]. These real experiments demonstrate that a text encoded into the DNA alphabet, then synthetized into DNA molecules, sequenced with the MinION, and finally processed using bioinformatics approaches can be fully recovered [28].

7.3.2. Identification of bacterial strains

Participants: Jacques Nicolas, Emeline Roux, Grégoire Siekaniec, Clara Delahaye.

Our aim is to provide rapid algorithms for the identification of bacteria at the finest taxonomic level. We have developed an expertise in the use of the MinIon long read technology and have produced and assembled many genomes for lactic bacteria in cooperation with INRA STLO, which have been made available on the Microscope platform at Genoscope (http://www.genoscope.cns.fr/agc/microscope/). We have developed a first classifier that demonstrates the possibility to identify isolated strains with spaced seed indexing of the noisy long reads produced by the MinIon.

7.4. Benchmarks and Reviews

7.4.1. Evaluation of error correction tools for long Reads

Participants: Lolita Lecompte, Pierre Peterlongo.

Long read technologies, such as Pacific Biosciences and Oxford Nanopore, have high error rates (from 9% to 30%). Hence, numerous error correction methods have been recently proposed, each based on different approaches and, thus, providing different results. As this is important to assess the correction stage for downstream analyses, we designed the ELECTOR software, providing evaluation of long read correction methods. This software generates additional quality metrics compared to previous existing tools. It also scales to very long reads and large datasets and is compatible with a wide range of state-of-the-art error correction tools [17]. ELECTOR is freely available at https://github.com/kamimrcht/ELECTOR.

7.4.2. Evaluation of insertion variant callers on real human data

Participants: Wesley Delage, Claire Lemaitre.

Insertion variants are one of the most common types of structural variation. Although such variants have many biological impacts on species evolution and health, they have been understudied because they are very difficult to detect with short read re-sequencing data. Recently, with the commercialization of novel long reads technologies, insertion variants are finally being discovered and referenced in human populations. Thanks to several international efforts, some gold standard call sets have been produced in 2019, referencing tens of thousands insertions. On these datasets, all existing short-read insertion variant callers, including our own method MindTheGap [9] which overtook others on simulated data, can reach at most 5 to 10 % of the referenced insertion variants. In this work, we propose a classification of the different types of insertion variants, based on the genomic context of the insertion site and the levels of duplication contained in the inserted sequence or within its breakpoints. In a detailed benchmark, we then analyze which of these types are the most impacted by the low recall of existing methods. Finally, by simulating various identified factors of difficulty, we investigate the causes of low recall and how these can be bypassed or improved in existing algorithms.

7.4.3. Modeling activities in cooperation with Inria project Dyliss

Participant: Jacques Nicolas.

J. Nicolas has maintained a partial activity with its previous research team Dyliss. In this framework, we have explored the use of Formal Concept Analysis (FCA) to ease the analysis of biological networks. The PhD thesis of L. Bourneuf on graph compression using FCA, defended this year, has introduced a new extension of FCA for this purpose, working on triplet concepts, which correspond to overlapping bicliques in graphs. The search space of concepts for graph compression has been presented in [21]. FCA applied to data on the steady states of a Boolean network and the dependencies between its proteins allowed to build a classifier used to analyze the states according to the phenotypic signatures of its network components. We have identified variants to the phenotypes and characterized hybrid phenotypes [19].

7.5. Bioinformatics Analysis

7.5.1. Genomics of Brassicaceae and agro-ecosystems insects

Participant: Fabrice Legeai.

Through its long term collaboration with INRA IGEPP, and its support to the BioInformatics of Agroecosystems Arthropods platform (http://bipaa.genouest.org), GenScale is involved in various genomic projects in the field of agricultural research. First, on plant genomics, we helped to identify duplicated copies of genes and repeated elements in the Brassica genomes [14]. Then, on major agricultural pests or their natural ennemies such as parasitoids, we conducted large scale analyses on the expression of effector genes involved in the adaptation of pea aphids to their host-plants [13]. Finally, we explored the expression of genes related to the virus machinery of bathyplectes parasitoids wasp of the alfalfa weevil [18].

7.5.2. Structural genome analysis of S. pyogenes strains

Participants: Emeline Roux, Dominique Lavenier.

The *S. pyogenes* bacteria is responsible for many human infections. With the increase in the prevalence of infections (750 million infections per year worldwide and 4th in terms of mortality from bacterial infection), a better understanding of adaptive and evolutionary mechanisms at play in this bacteria is essential. The molecular characterization of the different strains is done by the *emm* gene. A statistical analysis of the different types of *emm* on the Brittany population shows 3 main dynamics: sporadic types, endemic types or epidemic types. The last case was observed in Brittany for the type *emm75* between 2009 and 2017. Two hypotheses can be considered: (1) the emergence of a new subtype or winning clone in an unimmunized population; (2) increased pathogenicity through genetic evolution of the strains, including the acquisition of new virulence factors. In collaboration with the microbiology department of the Rennes Hospital, we sequenced more than 30 *S. pyogenesemm75* strains (Oxford Nanopore MinION sequencing) in order to study the dynamic of the epidemic thought their structural genomic variation.

7.5.3. Linking allele-specific expression and natural selection in wild populations

Participants: Mohammed Amin Madoui, Pierre Peterlongo.

Allele-specific expression (ASE) is now a widely studied mechanism at cell, tissue and organism levels. However, population-level ASE and its evolutionary impacts have still never been investigated. Here, we hypothesized a potential link between ASE and natural selection on the cosmopolitan copepod *Oithona similis*. We combined metagenomic and metatranscriptomic data from seven wild populations of the marine copepod *O. similis* sampled during the Tara Oceans expedition. We detected 587 single nucleotide variants (SNVs) under ASE and found a significant amount of 152 SNVs under ASE in at least one population and under selection across all the populations. This constitutes a first evidence that selection and ASE target more common loci than expected by chance, raising new questions about the nature of the evolutionary links between the two mechanisms [33].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Tank milk analysis

Participants: Dominique Lavenier, Jacques Nicolas.

The Seenergi company has developed a biotechnology protocol to detect cow mastitis directly by analyzing the DNA in the milk of the tanks. Cows are first genotyped. Since cows with mastitis produce a high level of lymphocytes, a DNA milk analysis can point out infested cows. Currently, DNA chips are used to support this analysis. We are currently investigating the possibility to use sequencing technologies in order to both reduce cost analysis and to extend the detection to larger herds.

8.2. Bilateral Grants with Industry

8.2.1. Rapsodyn project

Participants: Dominique Lavenier, Claire Lemaitre, Pierre Peterlongo, Gwendal Virlet.

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 to elaborate advanced tools dedicated to polymorphism detection and analysis.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Project Thermin: Differential characterization of strains of a bacterial species, Streptococcus thermophilus, with a Nanopore MinION

Participants: Jacques Nicolas, Emeline Roux, Grégoire Siekaniec, Dominique Lavenier.

Coordinator: J. Nicolas (Inria/Irisa, GenScale, Rennes)
Duration: 36 months (Oct. 2018 – Sept. 2021)
Partners: INRA (STLO, Agrocampus Rennes, E. Guédon and Y. Le Loir).

The Thermin project aims at exploring the capacities of a low cost third generation sequencing device, the Oxford Nanopore MinION, for rapid and robust pan-genome discrimination of bacterial strains and their phenotypes. It started with the recruitments of E. Roux (délégation Inria, Oct, 2018), a biochemist from Lorraine University, and G. Siekaniec (INRA -Inria collaboration, INRA grant), a new PhD student. We study pan-genomic representations of multiple genomes and the production of characteristic signatures of each genome in this context.

9.1.2. Project DNA-Store: Advanced error correction scheme for DNA-based data storage using nanopore technology

Participants: Dominique Lavenier, Emeline Roux.

Coordinator: L. Conde-Canencia (UBS, Lab-STCC, IAS)
Duration: 12 months (Feb. 2019 - Feb. 2020)
Partners: UBS (Lab-STCC, IAS, L. Conde-Canencia)

The DNA-Store project is funded by the Labex CominLabs. The goal is to explore the possibility to store information on DNA molecules. As DNA sequencing (the reading process) is performed with the Oxford Nanopore technology, powerful error correcting codes need to be developed together with dedicated genomic data processing.

9.2. National Initiatives

9.2.1. ANR

9.2.1.1. Project HydroGen: Metagenomics applied to ocean life study

Participants: Dominique Lavenier, Pierre Peterlongo, Claire Lemaitre.

Coordinator: P. Peterlongo (Inria/Irisa, GenScale, Rennes)
Duration: 42 months (Nov. 2014 – Apr. 2019)
Partners: CEA (Genoscope, 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 generated by the Tara Oceans expedition.

9.2.1.2. Project SpeCrep: speciation processes in butterflies

Participants: Dominique Lavenier, Fabrice Legeai, Claire Lemaitre, Pierre Peterlongo.

Coordinator: M. Elias (Museum National d'Histoire Naturelle, Institut de Systématique et d'Evolution de la Biodiversité, Paris)
Duration: 48 months (Jan. 2015 – Jul. 2019)
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.

9.2.1.3. Project Supergene: The consequences of supergene evolution.

Participants: Anne Guichard, Dominique Lavenier, Fabrice Legeai, Claire Lemaitre, Pierre Peterlongo.

Coordinator: M. Joron (Centre d'Ecologie Fonctionnelle et Evolutive (CEFE) UMR CNRS 5175, Montpellier)
Duration: 48 months (Nov. 2018 – Oct. 2022)
Partners: CEFE (Montpellier), MNHN (Paris), Genscale Inria/IRISA Rennes.

The Supergene project aims at better understanding the contributions of chromosomal rearrangements to adaptive evolution. Using the supergene locus controlling adaptive mimicry in a polymorphic butterfly from the Amazon basin (*H. numata*), the project will investigate the evolution of inversions involved in adaptive polymorphism and their consequences on population biology. GenScale's task is to develop new efficient methods for the detection and genotyping of inversion polymorphism with several types of re-sequencing data.

9.2.1.4. Project SeqDigger: Search engine for genomic sequencing data

Participants: Dominique Lavenier, Claire Lemaitre, Pierre Peterlongo.

Coordinator: P. Peterlongo
Duration: 48 months (jan. 2020 – Dec. 2024)
Partners: Genscale Inria/IRISA Rennes, CEA genoscpoe, MIO Marseille, Institut Pasteur Paris

https://www.cesgo.org/seqdigger/

The central objective of the SeqDigger project is to provide an ultra fast and user-friendly search engine that compares a query sequence, typically a read or a gene (or a small set of such sequences), against the exhaustive set of all available data corresponding to one or several large-scale metagenomic sequencing project(s), such as New York City metagenome, Human Microbiome Projects (HMP or MetaHIT), Tara Oceans project, Airborne Environment, etc. This would be the first ever occurrence of such a comprehensive tool, and would strongly benefit the scientific community, from environmental genomics to biomedicine.

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, Pierre Peterlongo, Gwendal Virlet.

Coordinator: N. Nesi (Inra, IGEPP, Rennes)
Duration: 99 months (2012-2020)
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 detection and their application to the rapeseed plant. (http://www.rapsodyn.fr)

9.2.3. Programs from research institutions

9.2.3.1. Inria Project Lab: Neuromarkers

Participants: Dominique Lavenier, Pierre Peterlongo, Claire Lemaitre, Céline Le Beguec, Téo Lemane.

Coordinator: O. Colliot (Inria, Aramis, Paris)
Duration: 4 years (2017-2020)
Partners: Inria (Aramis, Pasteur, Dyliss, GenScale, XPOP), ICM

The Neuromarkers IPL aims to design imaging bio-markers of neuro-degenerative diseases for clinical trials and study of their genetic associations. In this project, GenScale brings its expertise in the genomics field. More precisely, given a case-control population, a first step is to identify small genetic variations (SNPs, small indels) from their genomes. Then, using these variations together with brain images (also partitioned into case-control data sets), the challenge is to select variants that present potential correlation with brain images.

9.3. European Initiatives

9.3.1. Collaborations in European Programs, Except FP7 & H2020

Program: ITN (Initiative Training Network)

Project acronym: IGNITE

Project title: Comparative Genomics of Non-Model Invertebrates

Duration: 48 months (April 2018, March 2022)

Coordinator: Gert Woerheide

Partners: Ludwig-Maximilians-Universität München (Germany), Centro Interdisciplinar de Investigação Marinha e Ambiental (Portugal), European Molecular Biology Laboratory (Germany), Université Libre de Bruxelles (Belgium), University of Bergen (Norway), National University of Ireland Galway (Ireland), University of Bristol (United Kingdom), Heidelberg Institute for Theoretical Studies (Germany), Staatliche Naturwissenschaftliche Sammlungen Bayerns (Germany), INRA Rennes (France), University College London (UK), University of Zagreb (Croatia), Era7 Bioinformatics (Spain), Pensoft Publishers (Bulgaria), Queensland Museum (Australia), Inria, GenScale (France), Institut Pasteur (France), Leibniz Supercomputing Centre of the Bayerische Akademie der Wissenschaften (Germany), Alphabiotoxine (Belgium)

Abstract: Invertebrates, i.e., animals without a backbone, represent 95 per cent of animal diversity on earth but are a surprisingly underexplored reservoir of genetic resources. The content and architecture of their genomes remain poorly characterised, but such knowledge is needed to fully appreciate their evolutionary, ecological and socio-economic importance, as well as to leverage the benefits they can provide to human well-being, for example as a source for novel drugs and biomimetic materials. IGNITE will considerably enhance our knowledge and understanding of animal genome knowledge by generating and analyzing novel data from undersampled invertebrate lineages and by developing innovative new tools for high-quality genome assembly and analysis.

9.3.2. Collaborations with Major European Organizations

Partner : PHC RILA 2019, Bulgaria

Two years France-Bulgaria bilateral Partnership Hubert Curien (PHC) RILA 2019 (project code : 43196Q). The topic of this project is "Integer Programming Approaches for Long-Reads Genome Assembly". Start year: 2019.

9.4. International Initiatives

9.4.1. HipcoGen

Title: High-Performance Combinatorial Optimization for Computational Genomics

International Partner (Institution - Laboratory - Researcher):

Information Sciences group of Los Alamos National Laboratory (LANL), Los Alamos, NM 87544, USA. coordinator - Hristo Djidjev

Start year: 2017

See also: 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.

9.4.2. Inria International Partners

9.4.2.1. Informal International Partners

- Free University of Brussels, Belgium: Genome assembly [P. Perterlongo, D. Lavenier]

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- Visit of Hristo Djidjev from Los Alamos National Laboratory, USA, June 2019
- Visit of Alla Lapidus and Anto korobeynikov, Center for Algoritmic Biotechnology, St. Petersburg State University, Russia, October 2019

9.5.2. Visits to International Teams

9.5.2.1. Research Stays Abroad

- Visit of R. Andonov at Los Alamos National Laboratory, USA, from March 23 to April 30th, 2019.
- Visit of D. Lavenier at Los Alamos National Laboratory, USA, from May 13th to May 24th, 2019

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Selection

10.1.1.1. Chair of Conference Program Committees

- seqBIM 20019 : national meeting of the sequence algorithms GT seqBIM [C. Lemaitre]

10.1.1.2. Member of the Conference Program Committees

- JOBIM 2019: French symposium of Bioinformatics [C. Lemaitre]
- seqBIM 2019 : national meeting of the sequence algorithms GT seqBIM [C. Lemaitre, P. Peterlongo]
- BIBM 2019, IEEE International Conference on Bioinformatics and Biomedicine [D. Lavenier]
- BIOKDD 2019, 10th International Workshop on Biological Knowledge Discovery from Big Data [D. Lavenier]
- IPDPS 2019, 33rd IEEE International Parallel and Distributed Processing Symposium [D. Lavenier]
- 11th International Conference on Bioinformatics and Computational Biology (BICOB-2019) [R. Andonov]
- RCAM 2019 (Recent Computational Advances in Metagenomics) [P. Peterlongo]

10.1.1.3. Reviewer

- ISMB 2019 (Intelligent Systems for Molecular Biology) [W. Delage, C. Lemaitre]

10.1.2. Journal

10.1.2.1. Reviewer - Reviewing Activities

- Bioinformatics [P. Peterlongo, D. Lavenier]
- BMC Bioinformatics [D. Lavenier]
- BMC Genomics [D. Lavenier, F. Legeai]
- Briefing in Bioinformatics [D. Lavenier]
- Nature Scientific Reports [F. Legeai]
- Genome Biology [D. Lavenier]
- Genomics [D. Lavenier]
- Machine Learning [J. Nicolas]

10.1.3. Invited Talks

- (IB)2 Research Day - Bruxelles - Oct 2019- "Metagenomics-Comparisons and variants" [P. Peterlongo]

10.1.4. Leadership within the Scientific Community

- Members of the Scientific Advisory Board of the GDR BIM (National Research Group in Molecular Bioinformatics) [P. Peterlongo, C. Lemaitre]

- Animator of the Sequence Algorithms axis (seqBIM GT) of the BIM and IM GDRs (National Research Groups in Molecular Bioinformatics and Informatics and Mathematics respectively) [C. Lemaitre]
- Animator of the INRA Center for Computerized Information Treatment "BBRIC" [F. Legeai]

10.1.5. 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]
- Member of the Scientific Council of Agrocampus Ouest (Institute for life, food and horticultural sciences and landscaping) [J. Nicolas]
- Expert for the Elixir French Service Delivery Plan [P. Peterlongo]

10.1.6. Research Administration

- Member of the CoNRS, section 06, [D. Lavenier]
- Member of the CoNRS, section 51, [D. Lavenier]
- Corresponding member of COERLE (Inria Operational Committee for the assesment of Legal and Ethical risks). Participation to the ethical group of IFB (French Elixir node, Institut Français de Bioinformatique) [J. Nicolas]
- Member of the steering committee of the INRA BIPAA Platform (BioInformatics Platform for Agro-ecosystems Arthropods) [D. Lavenier]
- Member of the steering committee of The GenOuest Platform (Bioinformatics Resource Center BioGenOuest) [D. Lavenier]
- Scientific Advisor of The GenOuest Platform (Bioinformatics Resource Center BioGenOuest) [J. Nicolas]
- Representative of the environnemental axis of the IRISA UMR [C. Lemaitre]
- 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]

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence : C. Delahaye, G. Siekaniec, Python, 48 h, L1, Univ. Rennes 1, France.

Licence : R. Andonov, Graph Algorithms, 60h, L3, Univ. Rennes 1, France.

Licence : K. Da Silva, Algorithms and Complexity, 36 h, L3, ENSAI, Rennes, France.

Licence : W. Delage, Introduction to Biostatistics, 36 h, L2, Univ. Rennes 1, France.

Master : R. Andonov, S. Francois, Operational research, 82h, M1 Miage, Univ. Rennes 1, France.

Master : G. Siekaniec, Python, 21 h, M1, Univ. Rennes 1, France.

Master : K. Da Silva, Statistical learning, 22h, M1, Univ. Rennes 1, France.

Master : C. Lemaitre, P. Peterlongo, Algorithms on Sequences, 52h, M2, Univ. Rennes 1, France.

Master : C. Lemaitre, T. Lemane, Bioinformatics of Sequences, 40h, M1, Univ. Rennes 1, France.

Master : P. Peterlongo, Experimental Bioinformactics, 24h, M1, ENS Rennes, France.

Master : F. Legeai, RNA-Seq, Metagenomics and Variant discovery, 12h, M2, National Superior School Of Agronomy, Rennes, France.

Master : R. Andonov, Advanced Algorithmics, 25h, Univ. Rennes 1, France.

Master : D. Lavenier, Memory Efficient Algorithms for Big Data, Engineering School, ESIR, Rennes.

Master : W. Delage, Internship jury, 6 h, M1, Univ. Rennes 1, France

10.2.2. Supervision

PhD in progress : S. François, Combinatorial Optimization Approaches for Bioinformatics, 01/10/2016-30/09/2019, 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, De novo local assembly approaches for the detection of complex genomic variations in rare diseases, 01/10/2017, J. Thévenon and C. Lemaitre.

PhD in progress: K. da Silva, Metacatalogue : a new framework for intestinal microbiota sequencing data mining, 01/10/2018, M. Berland, N. Pons and P. Peterlongo.

PhD in progress: G. Siekaniec, Differential characterization of strains of bacterial species, 01/10/2018, E. Guédon, E. Roux and J. Nicolas.

PhD in progress: C. Delahaye, Robust interactive reconstruction of polyploid haplotypes, 01/10/2019, J. Nicolas

PhD in progress: T. Lemane, unbiased detection of neurodegenerative structural variants using k-mer matrices, 01/10/2019, P. Peterlongo

10.2.3. Juries

- *Member of Habilitation thesis jury*. S. Caboche, Univ. Lille [P. Peterlongo] R. Eyraud, Univ. Marseille [J. Nicolas, referee]
- *President of PhD thesis jury*. J. Wanza, University of Nice Sophia Antipolis [D. Lavenier], M. Guillemé, Univ. Rennes [J. Nicolas]
- *Referee of Ph-D thesis jury*. Kevin Gravouil, Univ Clermont Auvergne [P. Peterlongo]
- *Member of PhD thesis juries*. Lyam Baudry, University Paris-Sorbonne [C. Lemaitre], Pierre Morisse, Univ. Rouen [P. Peterlongo], Pierre Marijon, Univ Lille [P. Peterlongo], Franklin Delehelle, University of Toulouse [D. Lavenier], Lucas Bourneuf, Univ. Rennes [J. Nicolas].
- *Member of PhD thesis committee*. Chi Nguyen Lam, Univ. Brest [D. Lavenier], Benjamin Churcheward, Univ. Nantes [D. Lavenier], Victor Gaborit, Inserm Nantes [P. Peterlongo], Guillaume Gautreau CEA [P. Peterlongo], Sébastien François, Univ. Rennes [J. Nicolas], Mikaïl Demirdelen, Univ. Rennes [J. Nicolas], Hugo Talibart, Univ. Rennes [J. Nicolas].

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

- Member of the Interstice editorial board [P. Peterlongo]

10.3.2. Interventions

- In educational institutions : Participation to operation "A la découverte de la recherche" in high schools [P. Peterlongo]
- Short Movie "Moi, mon chien et les autres: à la recherche des transcrits perdus", presented at Sciences en Courts, a local contest of popularization short movies made by PhD students (http://sciences-en-courts.fr/) [G. Siekaniec].

10.3.3. Internal action

- co-organizer of Sciences en cour[t]s event, Nicomaque association (http://sciences-en-courts.fr/) [W. Delage]

10.3.4. Creation of media or tools for science outreach

- Popularization article in Interstices. “Analyser le génome des océans” https://interstices.info/analyser-les-genomes-des-oceans/ [P. Peterlongo]

11. Bibliography

Major publications by the team in recent years

[1] G. BENOIT, C. LEMAITRE, D. LAVENIER, E. DREZEN, T. DAYRIS, R. URICARU, G. RIZK.*Reference-free compression of high throughput sequencing data with a probabilistic de Bruijn graph*, in "BMC Bioinformatics", September 2015, vol. 16, n° 1 [*DOI :* 10.1186/S12859-015-0709-7], https://hal.inria.fr/hal-01214682

[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° 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. LAVENIER.*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] S. FRANÇOIS, 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, p. 1-11 [*DOI :* 10.1101/231324], https://hal.inria.fr/hal-01666830

[6] C. GUYOMAR, F. LEGEAI, E. JOUSSELIN, C. C. MOUGEL, C. LEMAITRE, J.-C. SIMON.*Multi-scale characterization of symbiont diversity in the pea aphid complex through metagenomic approaches*, in "Microbiome", December 2018, vol. 6, n° 1 [*DOI :* 10.1186/S40168-018-0562-9], https://hal.archives-ouvertes.fr/hal-01926402

[7] 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://hal.inria.fr/hal-01566246

[8] C. MARCHET.*From reads to transcripts: de novo methods for the analysis of transcriptome second and third generation sequencing*, Université de Rennes 1, September 2018, https://tel.archives-ouvertes.fr/tel-01939193

[9] 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° 24, p. 3451-3457 [*DOI :* 10.1093/BIOINFORMATICS/BTU545], https://hal.inria.fr/hal-01081089

[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

Articles in International Peer-Reviewed Journal

[11] R. ANDONOV, H. DJIDJEV, S. FRANÇOIS, D. LAVENIER.*Complete Assembly of Circular and Chloroplast Genomes Based on Global Optimization*, in "Journal of Bioinformatics and Computational Biology", 2019, p. 1-28, forthcoming [*DOI :* 10.1142/S0219720019500148], https://hal.archives-ouvertes.fr/hal-02151798

[12] G. BENOIT, M. MARIADASSOU, S. ROBIN, S. SCHBATH, P. PETERLONGO, C. LEMAITRE.*SimkaMin: fast and resource frugal de novo comparative metagenomics*, in "Bioinformatics", September 2019, p. 1-2 [*DOI :* 10.1093/BIOINFORMATICS/BTZ685], https://hal.inria.fr/hal-02308101

[13] H. BOULAIN, F. LEGEAI, J. JAQUIÉRY, E. GUY, S. MORLIERE, J.-C. SIMON, A. SUGIO.*Differential Expression of Candidate Salivary Effector Genes in Pea Aphid Biotypes With Distinct Host Plant Specificity*, in "Frontiers in Plant Science", October 2019, vol. 10, p. 1-12 [*DOI :* 10.3389/FPLS.2019.01301], https://hal.inria.fr/hal-02378266

[14] J. FERREIRA DE CARVALHO, J. LUCAS, G. DENIOT, C. FALENTIN, O. FILANGI, M. GILET, F. LEGEAI, M. LODE, J. MORICE, G. TROTOUX, J.-M. AURY, V. BARBE, J. KELLER, R. SNOWDON, Z. HE, F. DENOEUD, P. WINCKER, I. BANCROFT, A.-M. CHÈVRE, M. ROUSSEAU-GUEUTIN.*Cytonuclear interactions remain stable during allopolyploid evolution despite repeated whole-genome duplications in Brassica*, in "Plant Journal", February 2019, vol. 98, n^{o} 3, p. 434-447 [*DOI :* 10.1111/TPJ.14228], https://hal-univ-rennes1.archives-ouvertes.fr/hal-02019346

[15] L. GUILLOT, L. DELAGE, A. VIARI, Y. VANDENBROUCK, E. COM, A. A. RITTER, R. LAVIGNE, D. MARIE, P. PETERLONGO, P. POTIN, C. PINEAU.*Peptimapper: proteogenomics workflow for the expert annotation of eukaryotic genomes*, in "BMC Genomics", January 2019, vol. 20, n^{o} 1, 56 [*DOI :* 10.1186/S12864-019-5431-9], https://hal.inria.fr/hal-01987197

[16] A. LIMASSET, J.-F. FLOT, P. PETERLONGO.*Toward perfect reads: self-correction of short reads via mapping on de Bruijn graphs*, in "Bioinformatics", February 2019 [*DOI :* 10.1093/BIOINFORMATICS/BTZ102], https://hal.inria.fr/hal-02407243

[17] C. MARCHET, P. MORISSE, L. LECOMPTE, A. LEFEBVRE, T. LECROQ, P. PETERLONGO, A. LIMASSET.*ELECTOR: evaluator for long reads correction methods*, in "NAR Genomics and Bioinformatics", March 2020, vol. 2, n^{o} 1 [*DOI :* 10.1093/NARGAB/LQZ015], https://hal.inria.fr/hal-02371117

[18] S. ROBIN, M. RAVALLEC, M. FRAYSSINET, J. WHITFIELD, V. JOUAN, F. LEGEAI, A.-N. VOLKOFF.*Evidence for an ichnovirus machinery in parasitoids of coleopteran larvae*, in "Virus Research", 2019, vol. 263, p. 189-206 [*DOI :* 10.1016/J.VIRUSRES.2019.02.001], https://hal.archives-ouvertes.fr/hal-02059774

[19] M. WERY, O. DAMERON, J. NICOLAS, E. RÉMY, A. SIEGEL.*Formalizing and enriching phenotype signatures using Boolean networks*, in "Journal of Theoretical Biology", 2019, vol. 467, p. 66-79 [*DOI :* 10.1016/J.JTBI.2019.01.015], https://hal.inria.fr/hal-02018724

International Conferences with Proceedings

[20] J. N. ALANKO, H. BANNAI, B. CAZAUX, P. PETERLONGO, J. STOYE.*Finding all maximal perfect haplotype blocks in linear time*, in "WABI 2019 - Workshop on Algorithms in Bioinformatics", Niagara Falls, United States, ACM, September 2019, p. 1-9, https://hal.inria.fr/hal-02187246

[21] L. BOURNEUF, J. NICOLAS.*Concept Lattices as a Search Space for Graph Compression*, in "ICFCA 2019 - 15th International Conference on Formal Concept Analysis", Francfort, Germany, D. C. L. B. SERTKAYA (editor), ICFCA: International Conference on Formal Concept Analysis, Springer, May 2019, vol. 15th International Conference, n^{o} 15, p. 274-289 [*DOI :* 10.1007/978-3-030-21462-3_18], https://hal.inria.fr/hal-02399578

National Conferences with Proceeding

[22] C. GUYOMAR, W. DELAGE, F. LEGEAI, C. MOUGEL, J.-C. SIMON, C. LEMAITRE.*Reference-guided genome assembly in metagenomic samples*, in "JOBIM 2019 - Journées Ouvertes Biologie, Informatique et Mathématiques", Nantes, France, July 2019, p. 1-8, https://hal.inria.fr/hal-02308257

[23] L. LECOMPTE, P. PETERLONGO, D. LAVENIER, C. LEMAITRE.*Genotyping Structural Variations using Long Read data*, in "JOBIM 2019 - Journées Ouvertes Biologie, Informatique et Mathématiques", Nantes, France, July 2019, p. 1-8, https://hal.inria.fr/hal-02288091

Conferences without Proceedings

[24] L. LECOMPTE, P. PETERLONGO, D. LAVENIER, C. LEMAITRE.*Genotyping Structural Variations using Long Read Data*, in "HiTSeq 2019 - Conference on High Throughput Sequencing", Basel, Switzerland, July 2019, p. 1-3, https://hal.inria.fr/hal-02289484

[25] C. MARCHET, M. KERBIRIOU, A. LIMASSET.*Indexing De Bruijn graphs with minimizers*, in "Recomb seq", Whashinton, United States, April 2019 [*DOI :* 10.1101/546309], https://hal.archives-ouvertes.fr/hal-02435086

[26] P. MORISSE, C. MARCHET, A. LIMASSET, T. LECROQ, A. LEFEBVRE.*CONSENT: Scalable self-correction of long reads with multiple sequence alignment*, in "Recomb Seq", Washinton, France, May 2019, https://hal.archives-ouvertes.fr/hal-02435116

[27] H. S. PHAM, G. VIRLET, D. LAVENIER, A. TERMIER.*Statistically Significant Discriminative Patterns Searching*, in "DaWaK 2019 - 21st International Conference on Big Data Analytics and Knowledge Discovery", Linz, Austria, Springer, August 2019, p. 105-115 [*DOI :* 10.1007/978-3-030-27520-4_8], https://hal.archives-ouvertes.fr/hal-02190793

Other Publications

[28] A. BADUAL.*Stockage d'information sur ADN*, Université Rennes 1, June 2019, https://hal.inria.fr/hal-02401641

[29] L. CONDE-CANENCIA, B. HAMOUM, D. LAVENIER, E. ROUX.*Error Correction Schemes for DNA Storage with Nanopore Sequencing*, July 2019, JOBIM 2019 - Journées Ouvertes Biologie, Informatique et Mathématiques, Poster, https://hal.archives-ouvertes.fr/hal-02400744

[30] K. DA SILVA, N. PONS, M. BERLAND, F. PLAZA OÑATE, M. ALMEIDA, P. PETERLONGO.*From genomics to metagenomics: benchmark of variation graphs*, July 2019, JOBIM 2019 - Journées Ouvertes Biologie, Informatique et Mathématiques, Poster, https://hal.inria.fr/hal-02284559

[31] V. EPAIN.*De novo long reads assembly using integer linear programing*, Université de Rennes 1 [UR1], September 2019, https://hal.inria.fr/hal-02413832

[32] J. GAUTHIER, C. MOUDEN, T. SUCHAN, N. ALVAREZ, N. ARRIGO, C. RIOU, C. LEMAITRE, P. PETERLONGO.*DiscoSnp-RAD: de novo detection of small variants for population genomics*, October 2019, working paper or preprint, https://hal.inria.fr/hal-01634232

[33] R. LASO-JADART, K. SUGIER, E. PETIT, K. LABADIE, P. PETERLONGO, C. AMBROISE, P. WINCKER, J.-L. JAMET, M.-A. MADOUI.*Linking Allele-Specific Expression And Natural Selection In Wild Populations*, September 2019, working paper or preprint [*DOI :* 10.1101/599076], https://hal.inria.fr/hal-02275928

[34] D. LAVENIER, E. ROUX, L. CONDE-CANENCIA, B. HAMOUM.*Advanced Coding Schemes for DNA-Based Data Storage Using Nanopore Sequencing Technologies*, November 2019, Journées CominLabs 2019, Poster, https://hal.archives-ouvertes.fr/hal-02400656

[35] L. LECOMPTE, P. PETERLONGO, D. LAVENIER, C. LEMAITRE.*SVJedi : Structural variation genotyping using long reads*, July 2019, HiTSeq 2019 - Conference on High Throughput Sequencing, Poster, https://hal.inria.fr/hal-02290884

[36] T. LEMANE.*Search engine for genomic sequencing data*, Université Rennes1, July 2019, https://hal.inria.fr/hal-02410102

[37] K. THUILLIER.*Linear programming for metabolic network completion*, Inria Rennes - Bretagne Atlantique and University of Rennes 1, France, December 2019, https://hal.inria.fr/hal-02408003

References in notes

[38] R. S. HARRIS, P. MEDVEDEV.*Improved representation of sequence bloom trees*, in "Bioinformatics", 08 2019, btz662, https://doi.org/10.1093/bioinformatics/btz662

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

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
A5.1.1. - Engineering of interactive systems
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
A5.1.9. - User and perceptual studies
A5.5.4. - Animation
A5.6. - Virtual reality, augmented reality
A5.6.1. - Virtual reality
A5.6.2. - Augmented reality
A5.6.3. - Avatar simulation and embodiment
A5.6.4. - Multisensory feedback and interfaces
A5.10.5. - Robot interaction (with the environment, humans, other robots)
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
B5.6. - Robotic systems
B5.8. - Learning and training
B5.9. - Industrial maintenance

B6.4. - Internet of things
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
B9.4. - Sports
B9.6.6. - Archeology, History

1. Team, Visitors, External Collaborators

Research Scientists
Anatole Lécuyer [Team leader, Inria, Senior Researcher, HDR]
Ferran Argelaguet [Inria, Researcher]

Faculty Members
Bruno Arnaldi [INSA Rennes, Professor, HDR]
Valérie Gouranton [INSA Rennes, Associate Professor]
Maud Marchal [INSA Rennes, Associate Professor, HDR]

External Collaborators
Guillaume Moreau [Ecole Centrale de Nantes]
Jean-Marie Normand [Ecole Centrale de Nantes]
Mélanie Cogné [PhD, CHU Rennes]

Technical Staff
Hakim Si Mohammed [Inria, Engineer, from Dec 2019]
Alexandre Audinot [INSA Rennes, Engineer]
Adrien Reuzeau [INSA Rennes, Engineer]
Thierry Gaugry [INSA Rennes, Engineer]
Florian Nouviale [INSA Rennes, Research Engineer]
Ronan Gaugne [Université de Rennes 1, Research Engineer]
Kevin-Yoren Gaffary [INSA Rennes, Engineer, until Sep 2019]
Emeric Goga [SATT Ouest Valorisation, Engineer, until Mar 2019]
Carl-Johan Jorgensen [SATT Ouest Valorisation, Engineer, until Mar 2019]
Vincent Goupil [SATT Ouest Valorisation, Engineer, until Mar 2019]
Quentin Petit [INSA Rennes, Engineer, until Jul 2019]

PhD Students
Hakim Si Mohammed [Inria, until Dec 2019]
Guillaume Bataille [Orange Labs]
Antonin Bernardin [INSA Rennes]
Xavier de Tinguy de La Girouliere [ENS Paris]
Rebecca Fribourg [Inria]
Romain Lagneau [INSA Rennes]
Flavien Lécuyer [INSA Rennes]
Tiffany Luong [IRT B<>com]
Etienne Peillard [Inria]
Romain Terrier [IRT B<>com]
Mathis Fleury [Inria]
Guillaume Vailland [INSA Rennes]

Hugo Brument [Univ de Rennes I]
Diane Dewez [Inria]
Gerard Gallagher [Inria, until Nov 2019]
Victor Rodrigo Mercado Garcia [Inria]
Nicolas Olivier [InterDigital, from Feb 2019]
Gwendal Fouché [Inria, from Oct 2019]
Grégoire Richard [Inria, from Oct 2019]
Sebastian Vizcay [Inria, from Nov 2019]
Martin Guy [ECN, from Oct 2019]
Adelaide Genay [Inria, from Oct 2019]

Post-Doctoral Fellows

Giulia Lioi [Inria]
Thomas Howard [Inria]

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.

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 [44]: 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

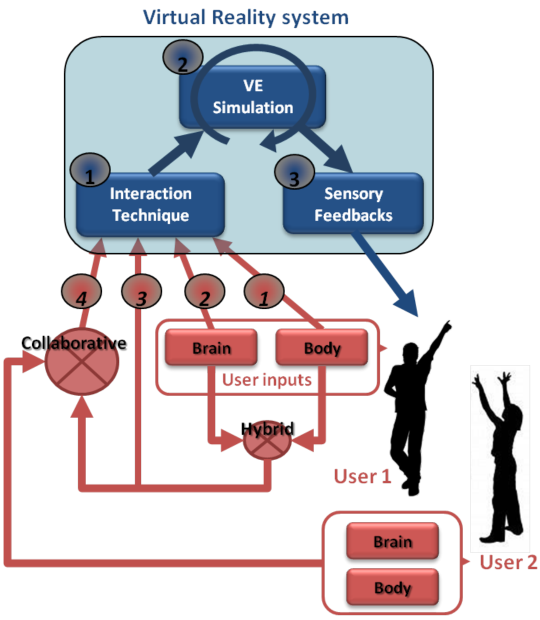


Figure 1. 3D hybrid interaction loop between one or multiple users and a virtual reality system. Top (in blue) three steps of 3D interaction with a virtual environment: (1-blue) interaction technique, (2-blue) simulation of the virtual environment, (3-blue) sensory feedbacks. Bottom (in red) different cases of interaction: (1-red) body-based, (2-red) mind-based, (3-red) hybrid, and (4-red) collaborative 3D interaction.

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 [45]. 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 movie industry, content customization, video games or attractions in theme parks,
- **Construction**: with virtual mock-ups design and review, or historical/architectural visits.
- **Cultural Heritage**: with acquisition, virtual excavation, virtual reconstruction and visualization

5. Highlights of the Year

5.1. Highlights of the Year

- Mélanie Cogné (Medical Doctor, PhD, CHU Rennes) has joined the Hybrid team as a new External Collaborator.
- Hybrid team has been strongly involved in the organization of the IEEE Virtual Reality Conference 2019 (IEEE VR), with F. Argelaguet (Program Chair) and A. Lécuyer (Panels Chair) and Jean-Marie Normand (Program Committee).
- The Immersia VR platform has celebrated its 20 years of existence at Inria Rennes/IRISA center, within the “20ans d’Immersia” event (November 2019).
- Hybrid team has organized a “VR Hackathon” at the Inria Rennes/IRISA Center, gathering around 20 participants (May 2019).

5.1.1. Awards

- IEEE VGTC Virtual Reality Technical Achievement Award 2019 was obtained by Anatole Lécuyer.
- IEEE VR Best 3DUI Contest Demo Award 2019: was obtained by Team Hybrid (Hugo Brument, Rebecca Fribourg, Gerard Gallagher, Thomas Howard, Flavien Lecuyer, Tiffany Luong, Victor Mercado, Etienne Peillard, Xavier de Tinguy, and Maud Marchal), for the demo entitled “Pyramid Escape: Design of Novel Passive Haptics Interactions for an Immersive and Modular Scenario” [11].

BEST PAPERS AWARDS :

[31]
E. PEILLARD, T. THEBAUD, J.-M. NORMAND, F. ARGELAGUET SANZ, G. MOREAU, A. LÉCUYER.*Virtual Objects Look Farther on the Sides: The Anisotropy of Distance Perception in Virtual Reality*, in "VR 2019 - 26th IEEE Conference on Virtual Reality and 3D User Interfaces", Osaka, Japan, IEEE, March 2019, p. 227-236 [*DOI :* 10.1109/VR.2019.8797826], https://hal.archives-ouvertes.fr/hal-02084069

[18]

R. GAUGNE, T. NICOLAS, Q. PETIT, M. OTSUKI, V. GOURANTON.*Evaluation of a Mixed Reality based Method for Archaeological Excavation Support*, in "ICAT-EGVE 2019 - International Conference on Artificial Reality and Telexistence - Eurographics Symposium on Virtual Environments", Tokyo, Japan, September 2019, p. 1-8, https://hal.inria.fr/hal-02272910

[23]

J. LACOCHE, T. DUVAL, B. ARNALDI, E. MAISEL, J. ROYAN.*Machine Learning Based Interaction Technique Selection For 3D User Interfaces*, in "EuroVR 2019 - 16th EuroVR International Conference", Tallinn, Estonia, Springer, October 2019, p. 33-51 [*DOI :* 10.1007/978-3-030-31908-3_3], https://hal.archives-ouvertes.fr/hal-02292434

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, Vincent Goupil, Carl-Johan Jorgensen, Emeric Goga, Adrien Reuzeau and Alexandre Audinot
- 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 enhanced Petri net and state machine models 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.

RELEASE FUNCTIONAL DESCRIPTION: Adding state machine handling for scenario description in addition to the already existing petri net format. Improved scenario editor

- Participants: Florian Nouviale, Valérie Gouranton, Bruno Arnaldi, Vincent Goupil, Emeric Goga, Carl-Johan Jorgensen, Adrien Reuzeau and Alexandre Audinot
- Contact: Valérie Gouranton
- Publications: Versatile Scenario Guidance for Collaborative Virtual Environments - Roles in Collaborative Virtual Environments for Training - 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 modules using a dedicated graphical language and a simple Graphical User Interface (GUI). This 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 60000 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: Cedric Riou, Thierry Gaugry, Anatole Lécuyer, Fabien Lotte, Jussi Tapio Lindgren, Laurent Bougrain, Maureen Clerc and Théodore Papadopoulo
- Partners: INSERM - GIPSA-Lab
- Contact: Anatole Lécuyer
- URL: http://openvibe.inria.fr

6.4. Platforms

6.4.1. Immerstar

- Participants: Florian Nouviale, Ronan Gaugne
- URL: http://www.irisa.fr/immersia/

With the two virtual reality technological platforms 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 was granted by an Inria funding for the 2015-2019 period which had enabled several important evolutions. In particular, in 2018, a haptic system covering the entire volume of the Immersia platform was installed, allowing various configurations from single haptic device usage to dual haptic devices usage with either one or two users. In addition, a motion platform designed to introduce motion feedback for powered wheelchair simulations has also been incorporated (see Figure 2).

We celebrated the twentieth anniversary of the Immersia platform in November 2019 by inaugurating the new haptic equipment. We proposed scientific presentations and received 150 participants, and visits for the support services in which we received 50 persons.

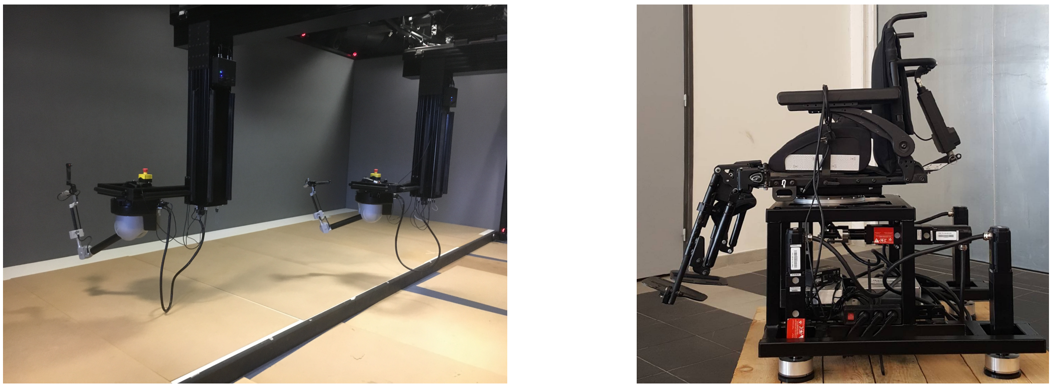


Figure 2. Immersia platform: (Left) "Scale-One" Haptic system for one or two users. (Right) Motion platform for a powered wheelchair simulation.

Figure 3. (Left) The LSI project in Immersia. (Right) The LSI project in Ars Electronica.

Immersia hosted and supported several projects and experiments during the year such as the Adapt [34], Sunset [2] or Introspect ([19], [29]) projects. Moreover, the platform was involved in the multidisciplinary project LSI led by the invited professor Franz Fischnaller from the Academy of Fine Arts of Torino. The project was implemented in Immersia with the support of the Immersia staff and deployed in the Ars Electronica center 3.

7. New Results

7.1. Virtual Reality Tools and Usages

7.1.1. Studying the Mental Effort in Virtual Versus Real Environments

Participants: Tiffany Luong, Ferran Argelaguet, Anatole Lécuyer [contact].

Is there an effect of Virtual Reality (VR) Head-Mounted Display (HMD) on the user's mental effort? In this work, we compare the mental effort in VR versus in real environments [26]. An experiment (N=27) was conducted to assess the effect of being immersed in a virtual environment (VE) using a HMD on the user's mental effort while performing a standardized cognitive task (the wellknown N-back task, with three levels of difficulty (1,2,3)). In addition to test the effect of the environment (i.e., virtual versus real), we also explored the impact of performing a dual task (i.e., sitting versus walking) in both environments on mental effort. The mental effort was assessed through self-reports, task performance, behavioural and physiological measures. In a nutshell, the analysis of all measurements revealed no significant effect of being immersed in the VE on the users' mental effort. In contrast, natural walking significantly increased the users' mental effort. Taken together, our results support the fact that there is no specific additional mental effort related to the immersion in a VE using a VR HMD.

7.1.2. Influence of Personality Traits and Body Awareness on the Sense of Embodiment in VR

Participants: Diane Dewez, Rebecca Fribourg, Ferran Argelaguet, Anatole Lécuyer [contact].

With the increasing use of avatars in virtual reality, it is important to identify the factors eliciting the sense of embodiment. This work reports an exploratory study aiming at identifying internal factors (personality traits and body awareness) that might cause either a resistance or a predisposition to feel a sense of embodiment towards a virtual avatar. To this purpose, we conducted an experiment (n=123) in which participants were immersed in a virtual environment and embodied in a gender-matched generic virtual avatar through a head-mounted display [16]. After an exposure phase in which they had to perform a number of visuomotor tasks, a virtual character entered the virtual scene and stabbed the participants' virtual hand with a knife (see Figure 4). The participants' sense of embodiment was measured, as well as several personality traits (Big Five traits and locus of control) and body awareness, to evaluate the influence of participants' personality on the acceptance of the virtual body. The major finding is that the locus of control is linked to several components of embodiment: the sense of agency is positively correlated with an internal locus of control and the sense of body ownership is positively correlated with an external locus of control. Taken together, our results suggest that the locus of control could be a good predictor of the sense of embodiment. Yet, further studies are required to confirm these results.
This work was done in collaboration with the MimeTIC team.

7.1.3. Consumer perceptions and purchase behavior of imperfect fruits and vegetables in VR

Participants: Jean-Marie Normand, Guillaume Moreau [contact].

This study investigates the effects of fruits and vegetables (FaVs) abnormality on consumer perceptions and purchasing behavior [9]. For the purposes of this study, a virtual grocery store was created with a fresh FaVs section, where 142 participants became immersed using an Oculus Rift DK2 Head-Mounted Display (HMD) software. Participants were presented either normal, slightly misshapen, moderately misshapen or severely misshapen FaVs. The study findings indicate that shoppers tend to purchase a similar number of FaVs whatever their level of deformity. However, perceptions of the appearance and quality of the FaVs depend on the degree of abnormality. Moderately misshapen FaVs are perceived as significantly better than those that are heavily misshapen but also "slightly" misshapen (except for the appearance of fruits).

Figure 4. From left to right: an example of a trajectory to draw during the experimental task; A view of the scene from behind; Another virtual character stabbing the participants' virtual hand at the end of the experiment to measure their response to the threat on their virtual body.

This work was done in collaboration with Audecia Recherce, the University of Reading and the University of Tokyo.

7.1.4. Am I better in VR with a real audience?

Participants: Romain Terrier, Valérie Gouranton [contact], Bruno Arnaldi.

We designed an experimental study to investigate the effects of a real audience on social inhibition [33]. The study is a virtual reality (VR) and multiuser application (see Figure 5). The experience is locally or remotely shared. The application engages one user and a real audience (i.e., local or remote conditions). A control condition is designed where the user is alone (i.e., alone condition). The objective performance (i.e., type and answering time) of users, when performing a categorization of numbers task in VR, is used to explore differences between conditions. In addition to this, the perceptions of others, the stress, the cognitive workload, and the presence of each user have been compared in relation to the location of the real audience. The results showed that in the presence of a real audience (in the local and remote conditions), user performance is affected by social inhibitions. Furthermore, users are even more influenced when the audience does not share the same room, despite others are less perceived.
This work was done in collaboration with IRT B COM.

Figure 5. Experimental setup for the social inhibition experiment in Virtual Reality.

7.1.5. Create by Doing – Action sequencing in VR

Participants: Flavien Lécuyer, Valérie Gouranton [contact], Adrien Reuzeau, Ronan Gaugne, Bruno Arnaldi.

In every virtual reality application, there are actions to perform, often in a logical order. This logical ordering can be a predefined sequence of actions, enriched with the representation of different possibilities, which we refer to as a scenario. Authoring such a scenario for virtual reality is still a difficult task, as it needs both the expertise from the domain expert and the developer. We propose [28] to let the domain expert create in virtual reality the scenario by herself without coding, through the paradigm of creating by doing (see Figure 6). The domain expert can run an application, record the sequence of actions as a scenario, and then reuse this scenario for other purposes, such as an automatic replay of the scenario by a virtual actor to check the obtained scenario, the injection of this scenario as a constraint or a guide for a trainee, or the monitoring of the scenario unfolding during a procedure.

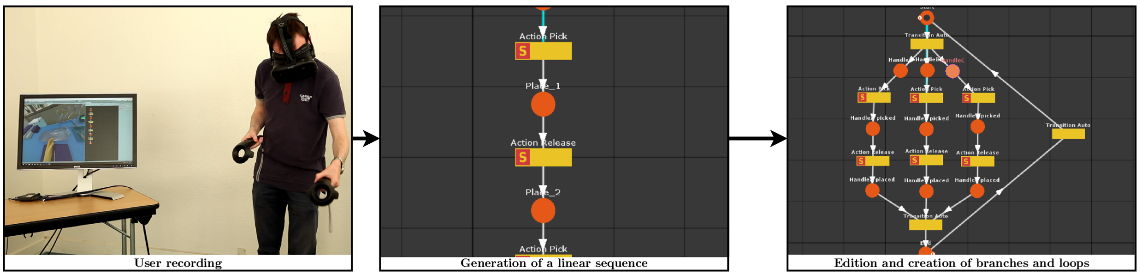


Figure 6. The proposed workflow for the creation of scenarios

7.1.6. Help! I Need a Remote Guide in my Mixed Reality Collaborative Environment

Participants: Valérie Gouranton [contact], Bruno Arnaldi.

The help of a remote expert in performing a maintenance task can be useful in many situations, and can save time as well as money. In this context, augmented reality (AR) technologies can improve remote guidance thanks to the direct overlay of 3D information onto the real world. Furthermore, virtual reality (VR) enables a remote expert to virtually share the place in which the physical maintenance is being carried out. In a traditional local collaboration, collaborators are face-to-face and are observing the same artifact, while being able to communicate verbally and use body language such as gaze direction or facial expression. These interpersonal communication cues are usually limited in remote collaborative maintenance scenarios, in which the agent uses an AR setup while the remote expert uses VR. Providing users with adapted interaction and awareness features to compensate for the lack of essential communication signals is therefore a real challenge for remote MR collaboration. However, this context offers new opportunities for augmenting collaborative abilities, such as sharing an identical point of view, which is not possible in real life. Based on the current task of the maintenance procedure, such as navigation to the correct location or physical manipulation, the remote expert may choose to freely control his/her own viewpoint of the distant workspace, or instead may need to share the viewpoint of the agent in order to better understand the current situation. In this work, we first focus on the navigation task, which is essential to complete the diagnostic phase and to begin the maintenance task in the correct location [8]. We then present a novel interaction paradigm, implemented in an early prototype, in which the guide can show the operator the manipulation gestures required to achieve a physical task that is necessary to perform the maintenance procedure. These concepts are evaluated, allowing us to provide guidelines for future systems targeting efficient remote collaboration in MR environments.

This work was done in collaboration with IRT B COM and UMR Lab-STICC, France.

7.1.7. Learning procedural skills with a VR simulator: An acceptability study

Participants: Valérie Gouranton [contact], Bruno Arnaldi.

Virtual Reality (VR) simulation has recently been developed and has improved surgical training. Most VR simulators focus on learning technical skills and few on procedural skills. Studies that evaluated VR simulators focused on feasibility, reliability or easiness of use, but few of them used a specific acceptability measurement tool. The aim of the study was to assess acceptability and usability of a new VR simulator for procedural skill training among scrub nurses, based on the Unified Theory of Acceptance and Use of Technology (UTAUT) model. The simulator training system was tested with a convenience sample of 16 non-expert users and 13 expert scrub nurses from the neurosurgery department of a French University Hospital. The scenario was designed to train scrub nurses in the preparation of the instrumentation table for a craniotomy in the operating room (OR). Acceptability of the VR simulator was demonstrated with no significant difference between expert scrub nurses and non-experts. There was no effect of age, gender or expertise. Workload, immersion and simulator sickness were also rated equally by all participants. Most participants stressed its pedagogical interest, fun and realism, but some of them also regretted its lack of visual comfort. This VR simulator designed to teach surgical procedures can be widely used as a tool in initial or vocational training [2], [43].
This work was achieved in collaboration with Univ. Rennes 2-LP3C, LTSI and the Hycomes team.

7.1.8. The Anisotropy of Distance Perception in VR

Participants: Etienne Peillard, Anatole Lécuyer, Ferran Argelaguet, Jean-Marie Normand, Guillaume Moreau [contact].

The topic of distance perception has been widely investigated in Virtual Reality (VR). However, the vast majority of previous work mainly focused on distance perception of objects placed in front of the observer. Then, what happens when the observer looks on the side? In this work, we study differences in distance estimation when comparing objects placed in front of the observer with objects placed on his side [31]. Through a series of four experiments (n=85), we assessed participants' distance estimation and ruled out potential biases. In particular, we considered the placement of visual stimuli in the field of view, users' exploration behavior as well as the presence of depth cues. For all experiments a two-alternative forced choice (2AFC) standardized psychophysical protocol was employed, in which the main task was to determine the stimuli that seemed to be the farthest one. In summary, our results showed that the orientation of virtual stimuli with respect to the user introduces a distance perception bias: objects placed on the sides are systematically perceived farther away than objects in front. In addition, we could observe that this bias increases along with the angle, and appears to be independent of both the position of the object in the field of view as well as the quality of the virtual scene. This work sheds a new light on one of the specificities of VR environments regarding the wider subject of visual space theory. Our study paves the way for future experiments evaluating the anisotropy of distance perception in real and virtual environments.

7.1.9. Study of Gaze and Body Segments Temporal Reorientation Behaviour in VR

Participants: Hugo Brument, Ferran Argelaguet [contact].

This work investigates whether the body anticipation synergies in real environments (REs) are preserved during navigation in virtual environments (VEs). Experimental studies related to the control of human locomotion in REs during curved trajectories report a top-down body segments reorientation strategy, with the reorientation of the gaze anticipating the reorientation of head, the shoulders and finally the global body motion [12]. This anticipation behavior provides a stable reference frame to the walker to control and reorient his/her body segments according to the future walking direction. To assess body anticipation during navigation in VEs, we conducted an experiment where participants, wearing a head-mounted display, performed a lemniscate trajectory in a virtual environment (VE) using five different navigation techniques, including walking, virtual steering (head, hand or torso steering) and passive navigation. For the purpose of this experiment, we designed a new control law based on the power-law relation between speed and curvature during human walking. Taken together, our results showed a similar ordered top-down sequence of reorientation of the gaze, head and shoulders during curved trajectories for all the evaluated techniques. However, the anticipation mechanism was significantly higher for the walking condition compared to the others. Finally, the results work pave the way to the better understanding of the underlying mechanisms of human navigation in VEs and to the design of navigation techniques more adapted to humans.

This work was done in collaboration with the MimeTIC team and the Interactive Media Systems Group (TU Wien, Vienna, Austria).

7.1.10. User-centered design of a multisensory power wheelchair simulator

Participants: Guillaume Vailland, Valérie Gouranton [contact].

Autonomy and social inclusion can reveal themselves everyday challenges for people experiencing mobility impairments. These people can benefit from technical aids such as power wheelchairs to access mobility and overcome social exclusion. However, power wheelchair driving is a challenging task which requires good visual, cognitive and visuo-spatial abilities. Besides, a power wheelchair can cause material damage or represent a danger of injury for others or oneself if not operated safely. Therefore, training and repeated practice are mandatory to acquire safe driving skills to obtain power wheelchair prescription from therapists. However, conventional training programs may reveal themselves insufficient for some people with severe impairments. In this context, Virtual Reality offers the opportunity to design innovative learning and training programs while providing realistic wheelchair driving experience within a virtual environment. In line with this, we propose a user-centered design of a multisensory power wheelchair simulator [34]. This simulator addresses classical virtual experience drawbacks such as cybersickness and sense of presence by combining 3D visual rendering, haptic feedback and motion cues. The simulator was showcased in the SOFMER conference [37].
This work has been done in collaboration with Rainbow team.

Figure 7. Wheelchair simulator.

7.1.11. Machine Learning Based Interaction Technique Selection For 3D User Interfaces

Participant: Bruno Arnaldi [contact].

A 3D user interface can be adapted in multiple ways according to each user's needs, skills and preferences. Such adaptation can consist in changing the user interface layout or its interaction techniques. Personalization systems which are based on user models can automatically determine the configuration of a 3D user interface in order to fit a particular user. In this work, we proposed to explore the use of machine learning in order to propose a 3D selection interaction technique adapted to a target user [23]. To do so, we built a dataset with 51 users on a simple selection application in which we recorded each user profile, his/her results to a

2D Fitts Law based pre-test and his/her preferences and performances on this application for three different interaction techniques. Our machine learning algorithm based on Support Vector Machines (SVMs) trained on this dataset proposes the most adapted interaction technique according to the user profile or his/her result to the 2D selection pre-test. Our results suggest the interest of our approach for personalizing a 3D user interface according to the target user but it would require a larger dataset in order to increase the confidence about the proposed adaptations.

7.1.12. The 3DUI Contest 2019

Participants: Hugo Brument, Rebecca Fribourg, Gerard Gallagher, Thomas Howard, Flavien Lécuyer, Tiffany Luong, Victor Mercado, Etienne Peillard, Xavier de Tinguy, Maud Marchal [contact].

Pyramid Escape: Design of Novel Passive Haptics Interactions for an Immersive and Modular Scenario
In this work, we present the design of ten different 3D user interactions using passive haptics and embedded in an escape game scenario in which users have to escape from a pyramid in a limited time [11]. Our solution is innovative by its modularity, allowing interactions with virtual objects using tangible props manipulated either directly using the hands and feet or indirectly through a single prop held in the hand, in order to perform several interactions with the virtual environment (VE). We also propose a navigation technique based on the "impossible spaces" design, allowing users to naturally walk through several overlapping rooms of the VE. All together, our different interaction techniques allow the users to solve several enigmas built into a challenging scenario inside a pyramid.

7.2. Augmented Reality Tools and Usages

7.2.1. Authoring AR by AR, abstraction and libraries

Participants: Flavien Lécuyer, Valérie Gouranton [contact], Adrien Reuzeau, Ronan Gaugne, Bruno Arnaldi.

The demand for augmented reality applications is rapidly growing. In many domains, we observe a new interest for this technology, stressing the need for more efficient ways of producing augmented content. Similarly to virtual reality, interactive objects in augmented reality are a powerful means to improve the experience. While it is now well democratized for virtual reality, interactivity is still finding its way into augmented reality. To open the way to this interactive augmented reality, we designed a new methodology for the management of the interactions in augmented reality, supported by an authoring tool for the use by designers and domain experts [27]. This tool makes the production of interactive augmented content faster, while being scalable to the needs of each application. Usually in the creation of applications, a large amount of time is spent through discussions between the designer (or the domain expert), carrying the needs of the application, and the developer, holding the knowledge to create it (see Figure 8). Thanks to our tool, we reduce this time by allowing the designer to create an interactive application, without having to write a single line of code.

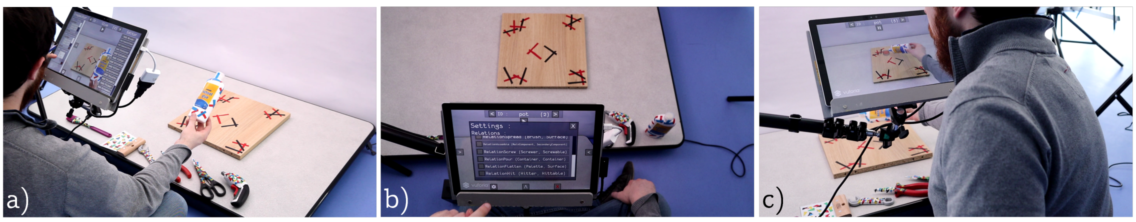


Figure 8. From left to right, the user (a) adds an interactive behaviour on a bottle of glue, (b) imports the interactions in the environment, and (c) uses the interaction to pour the virtual glue from the real bottle into a virtual pot

7.2.2. Studying Exocentric Distance Perception in Optical See-Through AR

Participants: Etienne Peillard, Ferran Argelaguet, Jean-Marie Normand, Anatole Lécuyer, Guillaume Moreau [contact].

While perceptual biases have been widely investigated in Virtual Reality (VR), very few studies have considered the challenging environment of Optical See-through Augmented Reality (OST-AR). Moreover, regarding distance perception, existing works mainly focus on the assessment of egocentric distance perception, i.e. distance between the observer and a real or a virtual object. In this work, we studied exocentric distance perception in AR, hereby considered as the distance between two objects, none of them being directly linked to the user. We report a user study (n=29) aiming at estimating distances between two objects lying in a frontoparallel plane at 2.1m from the observer (i.e. in the medium-field perceptual space). Four conditions were tested in our study: real objects on the left and on the right of the participant (called real-real), virtual objects on both sides (virtual-virtual), a real object on the left and a virtual one on the right (real-virtual) and finally a virtual object on the left and a real object on the right (virtual-real). Participants had to reproduce the distance between the objects by spreading two real identical objects presented in front of them (see Figure 9). The main findings of this study are the overestimation (20%) of exocentric distances for all tested conditions. Surprisingly, the real-real condition was significantly more overestimated (by about 4%,p=.0166) compared to the virtual-virtual condition, i.e. participants obtained better estimates of the exocentric distance for the virtual-virtual condition. Finally, for the virtual-real/real-virtual conditions, the analysis showed a non-symmetrical behavior, which suggests that the relationship between real and virtual objects with respect to the user might be affected by other external factors. Considered together, these unexpected results illustrate the need for additional experiments to better understand the perceptual phenomena involved in exocentric distance perception with real and virtual objects [30].

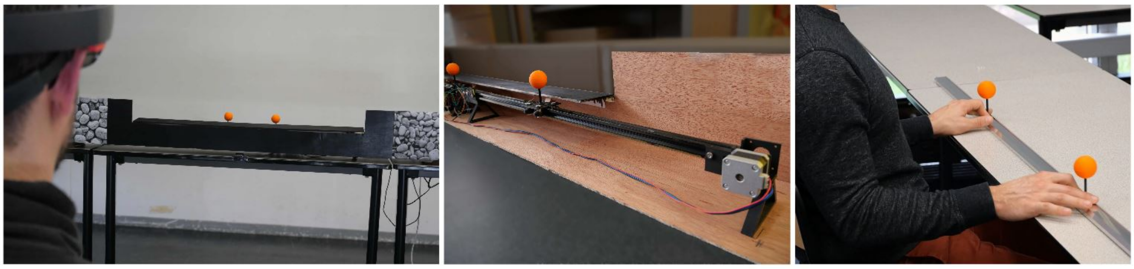


Figure 9. Left, bench displaying two real spheres. The hinge-actuated moving panel, opened here, could be automatically opened/closed to reveal/hide the visual stimuli. Center, one of the two rails of the bench, seen from behind. An orange sphere is attached on top of a trolley that can slide on the rail. The trolley is moved by a stepper motor through a belt. The other half of the bench is symmetrical. Right, participants could provide the perceived exocentric distance by placing two sliding spheres. After the participants placed the spheres the system automatically took a picture of both spheres which was used to measure the distance between both spheres.

7.2.3. Influence of virtual objects' shadows and lighting coherence in AR

Participants: Etienne Peillard, Jean-Marie Normand, Guillaume Moreau [contact].

This work focuses on how virtual objects' shadows as well as differences in alignment between virtual and real lighting influence distance perception in optical see-through (OST) augmented reality (AR) [5]. Four hypotheses are pro-posed: (H1) Participants underestimate distances in OST AR; (H2) Virtual objects' shadows improve distance judgment accuracy in OST AR; (H3) Shadows with different realism levels have different influence on distance perception in OST AR; (H4) Different levels of lighting misalignment between real and virtual lights have different influence on distance perception in OST AR scenes. Two experiments were designed with an OST head mounted display(HMD), the Microsoft HoloLens. Participants had to match the position of a virtual object displayed in the OST-HMD with a real target. Distance judgment accuracy was recorded under the different shadows and lighting conditions.The results validate hypotheses H2 and H4 but surprisingly showed no impact of the shape of virtual shadows on distance judgment accuracy thus rejecting hypothesis H3. Regarding hypothesis H1, we detected a trend toward underestimation; given the high variance of the data, more experiments are needed to confirm this result. Moreover, the study also reveals that perceived distance errors and completion time of trials increase along with targets' distance.

7.2.4. A study on differences in human perception in AR

Participants: Jean-Marie Normand, Guillaume Moreau [contact].

With the recent growth in the development of augmented reality (AR) technologies, it is becoming important to study human perception of AR scenes. In order to detect whether users will suffer more from visual and operator fatigue when watching virtual objects through optical see-through head-mounted displays (OST-HMDs), compared with watching real objects in the real world, we propose a comparative experiment including a virtual magic cube task and a real magic cube task [4]. The scores of the subjective questionnaires (SQ) and the values of the critical flicker frequency (CFF) were obtained from 18 participants. In our study, we use several electrooculogram (EOG) and heart rate variability (HRV) measures as objective indicators of visual and operator fatigue. Statistical analyses were performed to deal with the subjective and objective indicators in the two tasks. Our results suggest that participants were very likely to suffer more from visual and operator fatigue when watching virtual objects presented by the OST-HMD. In addition, the present study provides hints that HRV and EOG measures could be used to explore how visual and operator fatigue are induced by AR content. Finally, three novel HRV measures are proposed to be used as potential indicators of operator fatigue.

This work was done in collaboration with the Beijing Engineering Research Center of Mixed Reality and Advanced Display (School of Optics and Photonics, Beijing Institute of Technology, Beijing, China) and AICFVE (Beijing Film Academy, Beijing, China).

7.3. Physically-Based Simulation and Haptic Feedback

7.3.1. Design of haptic guides for pre-positioning assistance of a comanipulated needle

Participant: Maud Marchal [contact].

In minimally-invasive procedures like biopsy, the physician has to insert a needle into the tissues of a patient to reach a target. Currently, this task is mostly performed manually and under visual guidance. However, manual needle insertion can result in a large final positioning error of the tip that might lead to misdiagnosis and inadequate treatment. A way to solve this limitation is to use shared control; a gesture assistance paradigm that combines the cognitive skills of the operator with the precision, stamina and repeatability of a robotic or haptic device. In this paper, we propose to assist the physician with a haptic device that holds the needle and generates mechanical guides during the phase of manual needle pre-positioning. In the latter, the physician has to place the tip of the needle on a planned entry point, with a pre-defined angle of incidence. From this pre-operative information and also from intra-operative measurements, we propose to generate haptic cues, known as virtual fixtures, to guide the physician towards the desired position and orientation of the needle. It takes the form of five haptic guides, each one implementing virtual fixtures. We conducted a user study where those guides were compared to the unassisted reference gesture. The most constraining guide, in terms of assisted degrees of freedom, was highlighted as the one that provides the best results in terms of performance and user experience [20], [21].

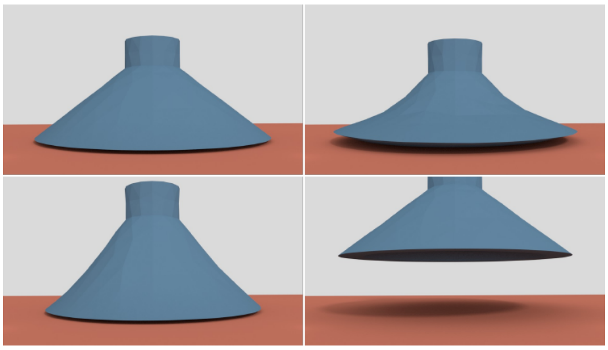
This work was done in collaboration with the Inria Rainbow team.

7.3.2. An Interactive Physically-based Model for Active Suction Phenomenon Simulation

Participants: Antonin Bernardin, Maud Marchal [contact].

While suction cups are widely used in Robotics, the literature is underdeveloped when it comes to the modelling and simulation of the suction phenomenon (see Figure 10). In this work, we present a novel physically-based approach to simulate the behavior of active suction cups. Our model relies on a novel formulation which assumes the pressure exerted on a suction cup during active control is based on constraint resolution. Our algorithmic implementation uses a classification process to handle the contacts during the suction phenomenon of the suction cup on a surface. Then, we formulate a convenient way for coupling the pressure constraint with the multiple contact constraints. We propose an evaluation of our approach through a comparison with real data, showing the ability of our model to reproduce the behavior of suction cups. Our approach paves the way for improving the design as well as the control of robotic actuators based on suction cups such as vacuum grippers.

This work was done in collaboration with the Inria Defrost team.



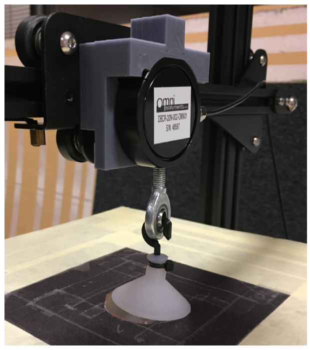


Figure 10. Left, Illustration of our constraint-based physically-based approach for simulating active suction cup phenomenon: (Top) the suction cup is actively stuck to the surface, (Bottom) is is then release until being completely in the air. Right, experimental setup for the force measurements. The suction cup is attached to a force sensor. When it is positioned on a flat surface, its cavity is linked to a vacuum pump with a regulator inbetween.

7.3.3. How different tangible and virtual objects can be while still feeling the same?

Participants: Xavier de Tinguy, Anatole Lécuyer, Maud Marchal [contact].

Tangible objects are used in Virtual Reality to provide human users with distributed haptic sensations when grasping virtual objects. To achieve a compelling illusion, there should be a good correspondence between the haptic features of the tangible object and those of the corresponding virtual one, i.e., what users see in the virtual environment should match as much as possible what they touch in the real world. This work [14] aims at quantifying how similar tangible and virtual objects need to be, in terms of haptic perception, to still feel the same. As it is often not possible to create tangible replicas of all the virtual objects in the scene, it is important to understand how different tangible and virtual objects can be without the user noticing (see Figure 11). This paper reports on the just-noticeable difference (JND) when grasping, with a thumb-index pinch, a tangible object which differ from a seen virtual one on three important haptic features: width, local orientation, and curvature. Results show JND values of 5.75%, 43.8%, and 66.66% of the reference shape for the width, local orientation, and local curvature features, respectively. These results will enable researchers in the field of Virtual Reality to use a reduced number of tangible objects to render multiple virtual ones.
This work was done in collaboration with the Inria Rainbow team.

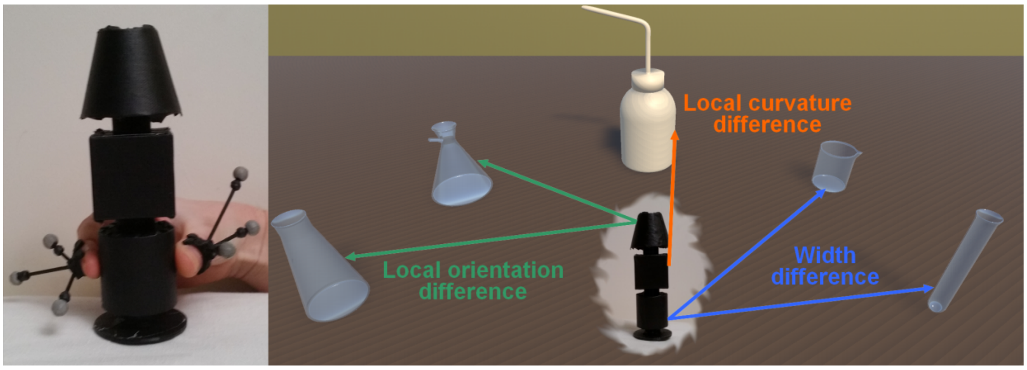


Figure 11. Understanding how different a tangible object (left) can be from virtual objects (right) without the user noticing the mismatch. We focused our study on three specific criteria: width, local orientation, and curvature.

7.3.4. Toward Universal Tangible Objects

Participants: Xavier de Tinguy, Maud Marchal, Anatole Lécuyer [contact].

Tangible objects are a simple yet effective way for providing haptic sensations in Virtual Reality. For achieving a compelling illusion, there should be a good correspondence between what users see in the virtual environment and what they touch in the real world. The haptic features of the tangible object should indeed match those of the corresponding virtual one in terms of, e.g., size, local shape, mass, texture. A straightforward solution is to create perfect tangible replicas of all the virtual objects in the scene. However, this is often neither feasible nor desirable. This work [15] presents an innovative approach enabling the use of few tangible objects to render many virtual ones (see Figure 12). The proposed algorithm analyzes the available tangible and virtual objects to find the best grasps in terms of matching haptic sensations. It starts by identifying several suitable pinching poses on the considered tangible and virtual objects. Then, for each pose, it evaluates a series of haptically-salient characteristics. Next, it identifies the two most similar pinching poses according to these metrics, one on the tangible and one on the virtual object. Finally, it highlights the chosen pinching pose, which provides the best matching sensation between what users see and touch. The effectiveness of our approach is evaluated through a user study. Results show that the algorithm is able to well combine several haptically-salient object features to find convincing pinches between the given tangible and virtual objects.

This work was done in collaboration with the Inria Rainbow team.

7.3.5. Investigating the recognition of local shapes using mid-air ultrasound haptics

Participants: Thomas Howard, Gerard Gallagher, Anatole Lécuyer, Maud Marchal [contact].

Mid-air haptics technologies are able to convey haptic sensations without any direct contact between the user and the haptic interface. One representative example of this technology is ultrasound haptics, which uses ultrasonic phased arrays to deliver haptic sensations. Research on ultrasound haptics is only in its beginnings, and the literature still lacks principled perception studies in this domain. This work [22] presents a series of human subject experiments investigating important perceptual aspects related to the rendering of 2D shapes by an ultrasound haptic interface (the Ultrahaptics STRATOS platform, see Figure 13). We carried out four user studies aiming at evaluating (i) the absolute detection threshold for a static focal point rendered via amplitude modulation, (ii) the absolute detection and identification thresholds for line patterns rendered via spatiotemporal modulation, (iii) the ability to discriminate different line orientations, and (iv) the ability to perceive virtual bumps and holes. These results shed light on the rendering capabilities and limitations of this novel technology for 2D shapes.

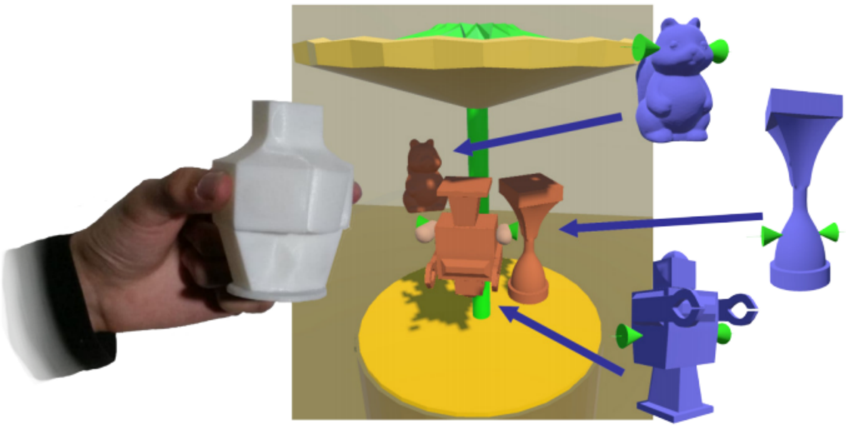


Figure 12. Illustration of our approach through a carousel of virtual objects that can be grasped using a single "universal" tangible object. The user is able to turn the virtual carousel and manipulate the three virtual objects using the suggested pinch poses (in green). These poses are proposed by our algorithm to best match the corresponding haptic pinching sensations on the tangible object.

This work was done in collaboration with the Inria Rainbow team.

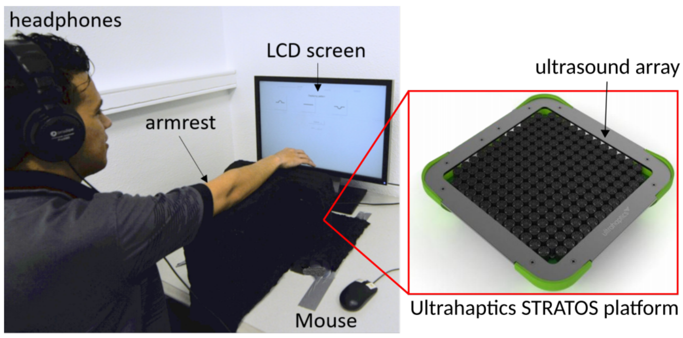


Figure 13. Experimental setup to investigate the recognition of local shapes using mid-air ultrasound haptics.

7.3.6. Touchy: Tactile Sensations on Touchscreens Using a Cursor and Visual Effects

Participants: Antoine Costes, Ferran Argelaguet, Anatole Lécuyer [contact].

Haptic enhancement of touchscreens usually involves vibrating motors that produce limited sensations or custom mechanical actuators that are difficult to widespread. In this work, we propose an alternative approach called "Touchy" to induce haptic sensations in touchscreens through purely visual effects [3]. Touchy introduces a symbolic cursor under the user's finger which shape and motion are altered in order to evoke haptic properties. This novel metaphor enables to address four different perceptual dimensions, namely: hardness, friction, fine roughness and macro roughness. Our metaphor comes with a set of seven visual effects that we compared with real texture samples within a user study conducted with 14 participants. Taken together

our results show that Touchy is able to elicit clear and distinct haptic properties: stiffness, roughness, reliefs, stickiness and slipperiness.
This work was achieved in collaboration with InterDigital.

7.3.7. *Investigating Tendon Vibration Illusions*

Participants: Salomé Lefranc [contact], Mélanie Cogné, Mathis Fleury, Anatole Lécuyer.

Illusion of movement induced by tendon vibration can be useful in applications such as rehabilitation of neurological impairments. In [40], we investigated whether a haptic proprioceptive illusion induced by a tendon vibration of the wrist congruent to the visual feedback of a moving hand could increase the overall illusion of movement. Tendon vibration was applied on the non-dominant wrist during 3 visual conditions: a moving virtual hand corresponding to the movement that the subjects could feel during the tendon vibration (Moving condition), a static virtual hand (Static condition), or no virtual hand at all (Hidden condition). There was a significant difference between the 3 visual feedback conditions, and the Moving condition was found to induce a higher intensity of illusion of movement and higher sensation of wrist's extension. Therefore, our study demonstrated the potentiation of illusion by visual cues congruent to the illusion of movement. Further steps will be to test the same hypothesis with stroke patients and use our results to develop EEG-based Neurofeedback including vibratory feedback to improve upper limb motor function after a stroke.
This work was achieved in collaboration with CHU Rennes and Inria EMPENN team.

7.4. Brain-Computer Interfaces

7.4.1. *Defining Brain-Computer Interfaces: A Human-Computer Interaction Perspective*

Participants: Hakim Si Mohammed, Ferran Argelaguet, Anatole Lécuyer [contact].

Regardless of the term used to designate them, Brain-Computer Interfaces (BCIs) are “Interfaces” between a user and a computer in the broad sense of the term. This paper aims to discuss how BCIs have been defined in the literature from the day the term was introduced by Jacques Vidal. In [32], from a Human-Computer Interaction perspective, we propose a new definition of Brain-Computer Interfaces as : "any artificial systems that directly converts brain activity into input of a computer process". As they are interfaces, such definition should not include the finality and objective of the system they are used to interact with. To illustrate this, we compared BCIs with other widely used Human-Computer Interfaces, and drew analogies in their conception and purpose.

This work was done in collaboration with the Inria LOKI team.

7.4.2. *A conceptual space for EEG-based brain-computer interfaces*

Participant: Anatole Lécuyer [contact].

Brain-Computer Interfaces have become more and more popular these last years. Researchers use this technology for several types of applications, including attention and workload measures but also for the direct control of objects by the means of BCIs. In [7] we present a first, multidimensional feature space for EEG-based BCI applications to help practitioners to characterize, compare and design systems, which use EEG-based BCIs. Our feature space contains 4 axes and 9 sub-axes and consists of 41 options in total as well as their different combinations. In addition we present the axes of our feature space and we position our feature space regarding the existing BCI and HCI taxonomies. We also showed how our work integrates the past works, and/or complements them.

7.4.3. *The use of haptic feedback in Brain-Computer Interfaces and Neurofeedback*

Participants: Mathis Fleury, Anatole Lécuyer [contact].

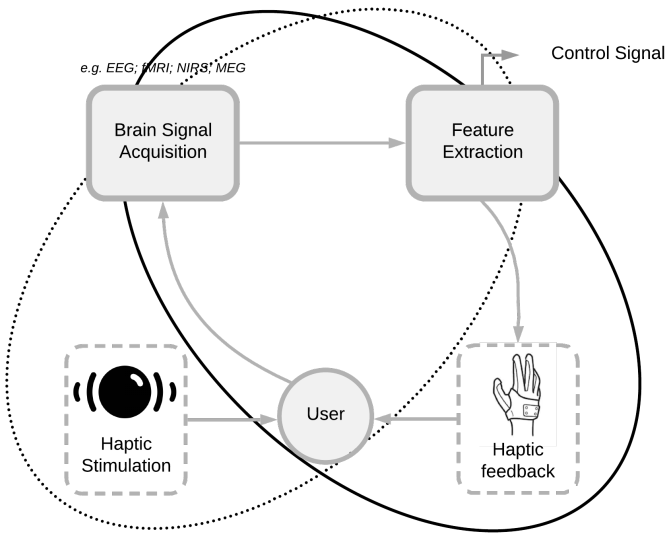


Figure 14. Using haptic feedback in active and reactive Brain-Computer Interfaces (BCI). In active BCI, haptics provide feedback from user's neural activity (black ellipse). In reactive BCI, haptics provide a stimulation to elicit a specific brain activity (black doted ellipse).

Neurofeedback (NF) and brain-computer interfaces are based on the recording of the cerebral activity associated with the requested task and the presentation of a feedback. The subject relies on the given feedback (visual, auditory or haptic) to learn and improve his mental strategy. It is therefore of crucial importance that it must be transmitted optimally. Historically, vision is the most used sensory modality in BCI/NF applications, but its use is raising potential issues. The more and more frequent use of haptic as a feedback modality reveals the limits of visual feedback; indeed, a visual feedback is not suitable in some cases, for individuals with an impaired visual system or during a mental motor imagery task (e.g. requiring a great abstraction). In such case, a haptic feedback would seem more appropriate. Haptic feedback has also been reported to be more engaging than visual feedback. This feedback could also contribute to close the sensory-motor loop. Haptic-based BCI/NF is a promising alternative for the design of the feedback and potentially improve the clinical efficacy of NF. In [38], [39] we have therefore surveyed the recent studies exploiting haptic feedback in BCI and NF.

This work was achieved in collaboration with the Inria EMPENN team.

7.4.4. *Efficacy of EEG-fMRI Neurofeedback for stroke rehabilitation: a pilot study*

Participants: Giulia Lioi, Mathis Fleury, Anatole Lécuyer [contact].

Recent studies have shown the potential of neurofeedback for motor rehabilitation after stroke. The majority of these NF approaches have relied solely on one imaging technique: mostly on EEG recordings. Recent study have gone further, revealing the potential of integrating complementary techniques such as EEG and fMRI to achieve a more specific regulation. In this exploratory work, multi-session bimodal EEG-fMRI NF for upper limb motor recovery was tested in four stroke patients.The feasibility of the NF training was investigated [41] with respect to the integrity of the cortico-spinal tract (CST), a well-established predictor of the potential for clinical improvement. Results indicated that patients exhibiting a high degree of integrity of the ipsilesional CST showed significant increased activation of the ipsilesional M1 at the end of the training. These preliminary

findings confirm the critical role of the CST integrity for stroke motor recovery and indicate that this is importantly related also to functional brain regulation of the ipsilesional motor cortex.
This work was achieved in collaboration with Inria EMPENN team.

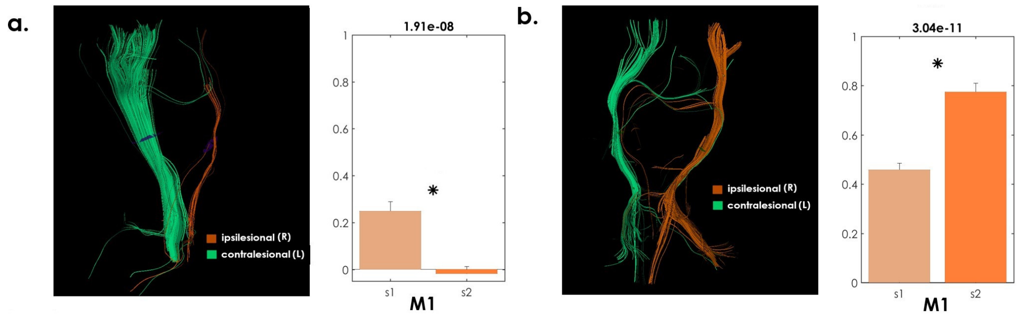


Figure 15. Example of CST reconstruction and primary motor cortex (M1) activation in two patients (a. and b.). Ipsilesional CST is plotted in orange and controlesional CST in green. The bar plot on the right hand side of the figure show the average (and standard error across NF training blocks) of BOLD contrast activation in the primary motor cortex in the first (s1) and second (s2) training session, with relative statistics.

7.4.5. A multi-target motor imagery training using EEG-fMRI Neurofeedback

Participants: Giulia Lioi, Mathis Fleury, Anatole Lécuyer [contact].

Upper limb recovery after stroke is a complex process. Recent studies have revealed the potential of neurofeedback training as an alternative or an aid to traditional therapies. Studies on cerebral plasticity and recovery after stroke indicate that premotor areas should be a preferred target for NF in the most severe patients while M1 stimulation may be more ef for patients with better recovery potential. Moreover, fMRI-NF studies (also on stroke patients) have shown that SMA is a robust correlate of motor imagery, while the activation of M1 is more dif to achieve, especially for short training sessions. Based on these results, in an exploratory work [13], we tested a dynamic NF training more strongly rewarding SMA activation in the NF training session and then increasing the M1 activation contribution in the NF session. We tested this novel approach on four stroke patients in a multisession bimodal EEG-fMRI NF training. To this end, we used an adaptive cortical region of interest (ROI) equal to a weighted combination of ipsilesional SMA and M1 activities and then varied the weights in order guide the patient training towards an improved activation of M1. Four chronic stroke patients with left hemiparesis participated to the study. The experimental protocol included an alternation of bimodal EEG-fMRI NF and unimodal EEG-only NF sessions. Preliminary results, on a short training duration, reveal the potential of a dynamic, multi-target/multimodal NF training approach.
This work was achieved in collaboration with Inria EMPENN team.

7.4.6. Bimodal EEG-fMRI Neurofeedback for upper motor limb rehabilitation

Participants: Giulia Lioi, Mathis Fleury, Anatole Lécuyer [contact].

There is a growing interest in Neurofeedback or Brain computer interfaces for stroke rehabilitation. Integrating EEG and fMRI, two highly complementary imaging modalities, has potential to provide a more specific and efficient stimulation of motor areas. In this exploratory work [25], we tested the feasibility of a multi-session EEG-fMRI NF protocol on four chronic stroke patients, and its potential for upper-limb recovery. All the patients were able to upregulate their activity during NF training with respect to rest in the ipsilesional SMA and M1. Three over four patients showed a significant increase in ipsilesional M1 activation at the end of the protocol. Of these three individuals, two exhibited an increase in FMA-UE score. Preliminary results from this pilot study showed feasibility of bimodal EEG-fMRI in chronic stroke patients and indicated the potential of this training protocol for upper-limb recovery.

This work was achieved in collaboration with Inria EMPENN team.

7.5. Cultural Heritage

7.5.1. Expressive potentials of motion capture in the Vis Insita musical performance

Participants: Ronan Gaugne [contact], Florian Nouviale, Valérie Gouranton.

The electronic music performance project Vis Insita [10] implements the design of experimental instrumental interfaces based on optical motion capture technology with passive infrared markers (MoCap), and the analysis of their use in a real scenic presentation context (Figure 16). Because of MoCap's predisposition to capture the movements of the body, a lot of research and musical applications in the performing arts concern dance or the sonification of gesture. For our research, we wanted to move away from the capture of the human body to analyse the possibilities of a kinetic object handled by a performer, both in terms of musical expression, but also in the broader context of a multimodal scenic interpretation.
This work was done in collaboration with Univ. Rennes 2, France.

Figure 16. The Vis Insita performance.

7.5.2. Interactive and Immersive Tools for Point Clouds in Archaeology

Participants: Ronan Gaugne [contact], Quentin Petit, Valérie Gouranton.

A framework is presented for an immersive and interactive 3D manipulation of large point clouds, in the context of an archaeological study [19]. The framework was designed in an interdisciplinary collaboration with archaeologists. We first applied this framework for the study of an 17th-century building of a Real Tennis court (Figure 17). We propose a display infrastructure associated with a set of tools that allows archaeologists to interact directly with the point cloud within their study process. The resulting framework allows an immersive navigation at scale 1:1 in a dense point cloud, the manipulation and production of cut plans and cross sections, and the positioning and visualisation of photographic views. We also apply the same framework to three other archaeological contexts with different purposes, a 13th century ruined chapel, a 19th-century wreck and a cremation urn from the Iron Age.
This work was done in collaboration with UMR CREAAH, France.

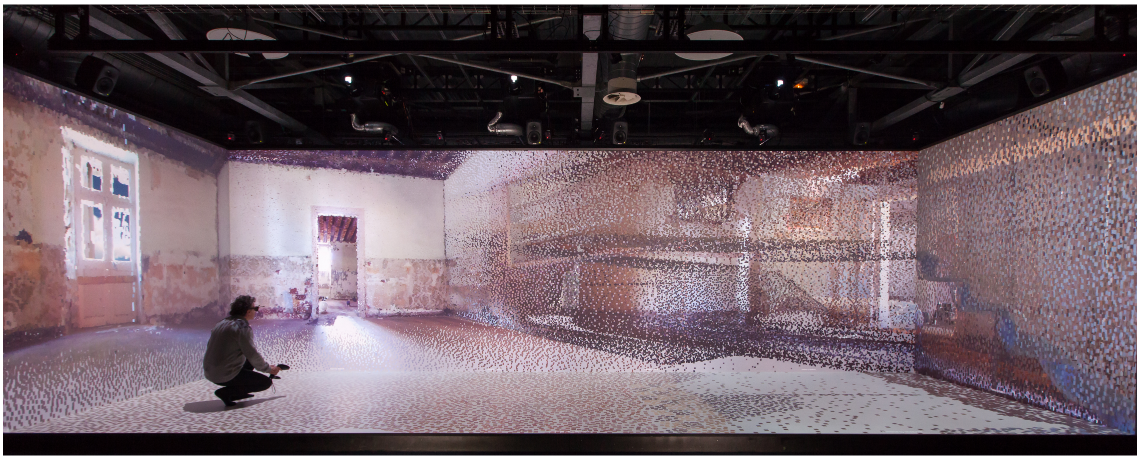


Figure 17. Immersive view of the point cloud of the Real Tennis building.

7.5.3. *Making virtual archeology great again (without scientific compromise)*

Participants: Ronan Gaugne, Valérie Gouranton [contact].

In the past two decades or so, digital tools have been slowly integrated as part of the archaeological process of information acquisition, analysis, and dissemination. We are now entering a new era, adding the missing piece to the puzzle in order to complete this digital revolution and take archaeology one step further into virtual reality (VR). The main focus of this work is the methodology of digital archaeology that fully integrates virtual reality, from beta testing to interdisciplinary teamwork. After data acquisition and processing necessary to construct the 3D model, we explore the analysis that can be conducted during and after the making or creation of the 3D environment and the dissemination of knowledge. We explain the relevance of this methodology through the case study on the intendant's palace, an 18th century archaeological site in Quebec City, Canada (Figure 18 left). With this experience, we believe that VR can prompt new questions that would never have occurred otherwise and can provide technical advantages in terms of gathering data in the same virtual space (Figure 18 right). We conclude that multidisciplinary input in archaeological research is once again proven essential in this new, inclusive and vast digital structure of possibilities [29].
This work was done in collaboration with UMR CREAAH, Inrap, France and Univ. Laval, Canada.

7.5.4. *Evaluation of a Mixed Reality based Method for Archaeological Excavation Support*

Participants: Ronan Gaugne [contact], Quentin Petit, Valérie Gouranton.

In the context of archaeology, most of the time, micro-excavation for the study of furniture (metal, ceramics...) or archaeological context (incineration, bulk sampling) is performed without complete knowledge of the internal content, with the risk of damaging nested artifacts during the process. The use of medical imaging coupled with digital 3D technologies, has led to significant breakthroughs by allowing to refine the reading of complex artifacts. However, archaeologists may have difficulties in constructing a mental image in 3 dimensions from the axial and longitudinal sections obtained during medical imaging, and in the same way to visualize and manipulate a complex 3D object on screen, and an inability to simultaneously manipulate and analyze a 3D image, and a real object. Thereby, if digital technologies allow a 3D visualization (stereoscopic screen, VR headset ...), they are not without limiting the natural, intuitive and direct 3D perception of the archaeologist on the material or context being studied. We therefore propose a visualization system based on optical see-through augmented reality that associates real visualization of archaeological material with data from medical imaging [18] (see Figure 19). This represents a relevant approach for composite or corroded

Figure 18. Left, model of the XVIIth century Palais de l'Intendant. Right, study of the reconstitution of the Palais de L'intendant and its neighborhood inside Immersia

objects or contexts associating several objects such as cremations. The results presented in the paper identify adequate visualization modalities to allow archaeologist to estimate, with an acceptable error, the position of an internal element in a particular archaeological material, an Iron-Age cremation block inside a urn.
This work was done in collaboration with Inrap, France and AIST (National Institute of Advanced Industrial Science and Technology), Japan.

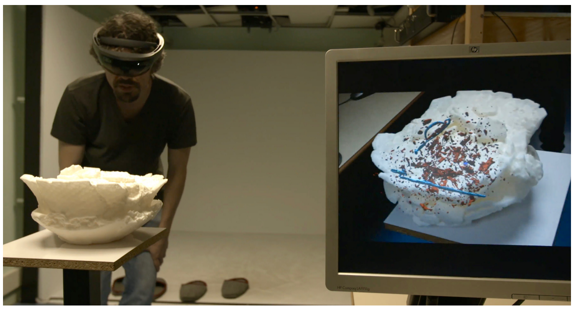


Figure 19. Evaluation of the mixed reality system

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Mensia Technologies

Participant: Anatole Lécuyer.

Mensia Technologies was an Inria start-up company created in November 2012 as a spin-off of Hybrid team. Mensia was focused on wellness and healthcare applications emerging from the BCI and Neurofeedback technologies. The Mensia startup benefited from the team's expertise and of valuable and proprietary BCI research results. Mensia was 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.

The contract between Hybrid and Mensia started in November 2013 and ended in August 2019 with the closing of the company. The contract supported the transfer of several softwares designed by Hybrid team (eg, OpenViBE and StateFinder) to Mensia Technologies for medical and multimedia applications of Mensia.

8.1.2. Orange Labs

Participants: Anatole Lécuyer [contact], Hakim Si-Mohammed, Ferran Argelaguet.

This four months contract between Hybrid and Orange labs (Jan - April 2019) covered the design of a proof of concept of a smart home system controlled using a brain computer interface in and augmented reality context.

8.2. Bilateral Grants with Industry

8.2.1. Orange Labs

Participants: Guillaume Bataille, Bruno Arnaldi, Valérie Gouranton [contact].

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".

In the context of this collaboration the following patent has been filled:

- "Dispositif d'affichage portatif de contenu 3D, système et procédé correspondants" (FR1914557), Guillaume Bataille, Bruno Arnaldi, Valérie Gouranton, Jérémy Lacoche. Filed in Dec. 2019.

8.2.2. InterDigital

Participants: Nicolas Olivier, Ferran Argelaguet, Anatole Lécuyer [contact].

This grant started in February 2019. It supports Nicolas's Olivier CIFRE PhD program with InterDigital company on "Avatar Stilization". This PhD is co-supervised with the MimeTIC team.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Labex Cominlabs SUNSET

Participants: Bruno Arnaldi, Valérie Gouranton [contact], Alexandre Audinot, Adrien Reuzeau.

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 RobotX

Participants: Bruno Arnaldi, Valérie Gouranton [contact], Alexandre Audinot.

RobotX (ROBOT for Intelligent Collaborative Surgery) is a one year Labex Cominlabs project (2019). The partners are MediCIS team from LTSI (INSERM and University of Rennes 1), Hybrid, Rainbow and Hycomes teams from IRISA and Inria Rennes, LP3C Lab - University Rennes 2, REV, ROMAS and PACCE teams from LS2N - Nantes, CHU Rennes, CHU Nantes, ICO (Institut de Cancerologie de l'Ouest). The objective of this exploratory action RobotX was to explore this issue and study initial feasibility of some methodological solutions. The long-term is to develop a new generation of intelligent and collaborative safe surgical robots.

Our contribution in the project was to study the development of Virtual Reality based simulated environments for surgical robotic systems for helping designing, evaluation and training of such systems. The objective was also to evaluate simulations of both technical and non technical. We developed a prototype of the Da Vinci robot with an haptic interface and different simulated tasks. We second studied the relevance of the software environments (#5 and #7) developed in previous projects (S3PM and SUNSET). We set up interactions by adding #5 semantics, which allow the robot arm to pick up objects. We also implemented a "Pick and place" exercise. A #7 scenario has been added to manage the user's actions and know when the exercise is over (Fig. 20).

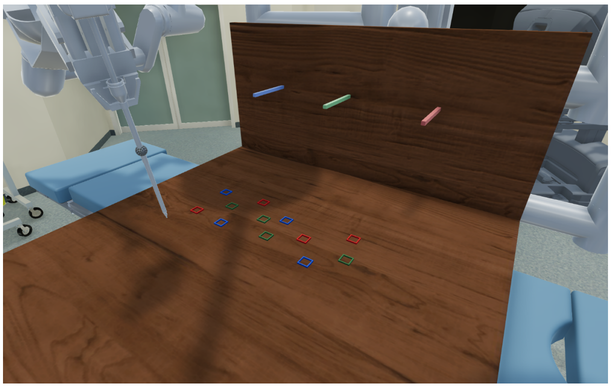


Figure 20. Pick and place exercise using #5 and #7 software

9.1.3. Labex Cominlabs HEMISFER

Participants: Mathis Fleury, Anatole Lécuyer [contact], Giulia Lioi.

HEMISFER is a 6-year project (2013-2019) 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. IRT b<>com

Participants: Ferran Argelaguet, Bruno Arnaldi [contact], Valérie Gouranton, Anatole Lécuyer, Maud Marchal, Florian Nouviale.

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 been regularly involved in collaborations with b<>com within various 3-year projects, such as ImData (on Immersive Interaction) and GestChir (on Augmented Healthcare) which both ended in 2016. Follow-up projects called NeedleWare (on Augmented Healthcare) and VUXIA (on Human Factors) have started respectively in 2016 and 2018.

9.1.5. CNPAO Project

Participants: Valérie Gouranton [contact], 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).

9.2. National Initiatives

9.2.1. ANR

9.2.1.1. ANR LOBBY-BOT

Participants: Anatole Lécuyer [contact], Maud Marchal, Victor Mercado.

LOBBY-BOT is a 4-year project (2017-2021) funded by the French National Research Agency (ANR). The objective of LOBBY-BOT is to address the scientific challenges of encountered-type haptic devices (ETHD), which are an alternative category of haptic devices relying on a mobile physical prop, usually actuated by a robot, that constantly follows the user hand, and encounter it only when needed. The project follows two research axes: a first one dealing with robot control, and the second one dealing with interaction techniques adapted to ETHD. The involvement of Hybrid relates to the second research axis of the project. The final project prototype will be used to assess the benefits of ETHD when used in an industrial use-case : the perceived quality in an automotive interior.

9.2.2. Inria projects

9.2.2.1. IPL BCI-LIFT

Participants: Anatole Lécuyer [contact], Hakim Si Mohammed.

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, Loki, 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.

9.2.2.2. IPL AVATAR

Participants: Anatole Lécuyer [contact], Ferran Argelaguet, Diane Dewez, Rebecca Fribourg.

AVATAR is a 4-year "Inria Project Lab" initiative (2018-2022) funded by Inria for supporting a national research effort on Avatars and Virtual Embodiment. This joint lab involves several Inria teams: Hybrid, Potioc, Loki, Mimetic, Graphdeco, Morpheo; as well as external partners: Univ. Bacelona, Faurecia and Technicolor companies. This project aims at improving several aspects of Avatars in immersive applications: reconstruction, animation, rendering, interaction, multi-sensory feedback, etc.

9.2.2.3. IPL NAVISCOPE

Participants: Ferran Argelaguet [contact], Gwendal Fouché.

NAVISCOPE is a 4-year "Inria Project Lab" initiative (2018-2022) funded by Inria for supporting a national research effort on image-guided navigation and visualization of large data sets in live cell imaging and microscopy. This joint lab involves several Inria teams: Serpico, Aviz, Beagle, Hybrid, Mosaic, Parietal, Morpheme; as well as external partners: INRA and Institute Curie. This project aims at improving visualization and machine learning methods in order to provide systems capable to assist the scientist to obtain a better understanding of massive amounts of information.

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

9.3.1.1. IMAGINE

Participants: Maud Marchal [contact], Thierry Gaugry, Romain Lagneau, Antonin Bernardin.

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

Abstract: 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.

9.3.1.2. H-REALITY

Participants: Anatole Lécuyer, Maud Marchal [contact], Thomas Howard, Gerard Gallagher.

Title: H-REALITY

Programm: H2020 - Fet Open

Duration: 2018 - 2021

Coordinator: Univ. Birmingham (UK)

Partners:

Univ. Birmingham (UK)

CNRS (France),

TU Delft (Netherlands),

ACTRONIKA (France),

ULTRAHAPTICS (UK)

Inria contact: Maud Marchal

Abstract: The vision of H-REALITY is to be the first to imbue virtual objects with a physical presence, providing a revolutionary, untethered, virtual-haptic reality: H-Reality. This ambition will be achieved by integrating the commercial pioneers of ultrasonic "non-contact" haptics, state-of-the-art vibrotactile actuators, novel mathematical and tribological modelling of the skin and mechanics of touch, and experts in the psychophysical rendering of sensation. The result will be a sensory experience where digital 3D shapes and textures are made manifest in real space via modulated, focused, ultrasound, ready for the unteathered hand to feel, where next-generation wearable haptic rings provide directional vibrotactile stimulation, informing users of an object's dynamics, and where computational renderings of specific materials can be distinguished via their surface properties. The implications of this technology will transform online interactions; dangerous machinery will be operated virtually from the safety of the home, and surgeons will hone their skills on thin air.

9.3.1.3. TACTILITY

Participants: Ferran Argelaguet [contact], Anatole Lécuyer, Maud Marchal, Sebastian Vizcay.

Title: Tactility

Programm: H2020 - ICT 25

Duration: July 2019 - June 2022

Coordinator: Fundación Tecnalia Research and Innovation (Spain)

Partners:

Aalborg University (Netherlands)

Universita Degli Studi di Genova (Itali),

Tecnalia Servia (Servia),

Universitat de Valencia (Spain),

Manus Machinae B.V. (Netherlands),

Smartex S.R.L (Italy),

Immersion (France)

Inria contact: Ferran Argelaguet

Abstract: TACTILITY is a multidisciplinary innovation and research action with the overall aim of including rich and meaningful tactile information into the novel interaction systems through technology for closed-loop tactile interaction with virtual environments. By mimicking the characteristics of the natural tactile feedback, it will substantially increase the quality of immersive VR experience used locally or remotely (tele-manipulation). The approach is based on transcutaneous electro-tactile stimulation delivered through electrical pulses with high resolution spatio-temporal distribution. To achieve it, significant development of technologies for transcutaneous stimulation, textile-based multi-pad electrodes and tactile sensation electronic skin, coupled with ground-breaking research of perception of elicited tactile sensations in VR, is needed. The key novelty is in the combination of: 1) the ground-breaking research of perception of electrotactile stimuli for the identification of the stimulation parameters and methods that evoke natural like tactile sensations, 2) the advanced hardware, that will integrate the novel high-resolution electrotactile stimulation system and state of the art artificial electronic skin patches with smart textile technologies and VR control devices in a wearable mobile system, and 3) the novel firmware, that handles real-time encoding and transmission of tactile information from virtual objects in VR, as well as from the distant tactile sensors (artificial skins) placed on robotic or human hands. Proposed research and innovation action would result in a next generation of interactive systems with higher quality experience for both local and remote (e.g., tele-manipulation) applications. Ultimately, TACTILITY will enable high fidelity experience through low-cost, user friendly, wearable and mobile technology.

9.3.1.4. Interreg ADAPT

Participants: Valérie Gouranton [contact], Bruno Arnaldi, Ronan Gaugne, Florian Nouviale, Yoren Gaffary, Alexandre Audinot.

Program: 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).

Inria contact: Valérie Gouranton

Abstract: The ADAPT 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.

Collaboration with Rainbow team.

9.4. International Initiatives

9.4.1. Informal International Partners

- Dr. Takuji Narumi and Prof. Michitaka Hirose from University of Tokyo (Japan), on "Virtual Embodiment"
- Dr. Hannes Kaufmann from Technical University Wien (Austria), on "3D Navigation in Virtual Environments"

- Prof. Reinhold Scherer from Graz University (Austria), on "Brain-Computer Interfaces and Augmented Reality"
- Prof. Jose Millan from Ecole Polytechnique Fédérale de Lausanne (Switzerland), on "Brain-Computer Interfaces and Sports"
- Dr. Mai Otsuki from AIST (Japan) on "Mixed Reality for Cultural Heritage"
- Dr. Karina Rodriguez Echavarria from University of Brighton (UK) on "Mixed Reality for Cultural Heritage"
- Prof. Franz Fischnaller from Albertina Academia of Fine Art of Torino (Italy) on "Immersive Art"
- Dr. Yuta Itoh from Tokyo Institute of Technology (Japan) on "Perception in Augmented Reality"

9.4.2. Participation in Other International Programs

9.4.2.1. ANR-FRQSC INTROSPECT

Participants: Valérie Gouranton [contact], Bruno Arnaldi, Ronan Gaugne, Flavien Lécuyer, Adrien Reuzeau.

INTROSPECT is a 3-year project funded by French ANR and "Fonds de Recherche Société et Culture" (FRQSC) from Quebec region, Canada. This international 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 also partners in the project to spread the results among the general public.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- Visit from Yutaro Hirao, Master Student at University of Tokyo (topic: “Virtual Embodiment”). Feb. 2019.
- Visit from Felix Putze, Researcher at University of Bremen (topic: “BCI and AR”). Feb. 2019.
- Visit from Franz Fischnaller, Professor at Academia of Fine Arts Albertina, Torino, Italy (topic: “Cultural Heritage”). From Jun. until Jul. 2019
- Visit from Marie-Anne Paradis, Master Student at University Laval, Québec, Canada (topic: “Cultural Heritage”). From Sept. 2018 until Mar. 2019.
- Visit from Nadia Zenati, Researcher at CDTA, Algeria (topic: “VR and AR”). Oct. 2019

9.5.2. Visits to International Teams

- Jean-Marie Normand spent 2 weeks (1 week in July 2019 and 1 week in September 2019) in the Augmented Vision Laboratory, Tokyo Institute of Technology, Tokyo, Japan.
- Valérie Gouranton and Ronan Gaugne spent 2 weeks, in May 2019, in the Eau-Terre-Environment laboratory of INRS, Québec, Canada where they presented INTROSPECT results in the GMPCA conference organized by the University of Montreal, and in the days of the Canadian Association of Archaeology, organized by the University Laval of Québec.

9.5.2.1. Research Stays Abroad

- Etienne Peillard spent 4 months (from June to October 2019) in the Augmented Vision Laboratory, Tokyo Institute of Technology, Tokyo, Japan.
- Flavien Lécuyer spent 3 months (From May to August 2019) in Vision and Numeric Systems Laboratory (LVSN), Québec, Canada.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

- Valérie Gouranton was Main Organizer of the Seminar "Art/Heritage/Culture" (IRISA, Rennes, February 2019), and Member of Organizing Committee of "Journée Science et Musique" (IRISA, Rennes, October 2019).
- Ronan Gaugne was Member of Organizing Committee of "Journée Science et Musique" (IRISA, Rennes, October 2019)
- Florian Nouviale was Member of Organizing Committee of the "Journée Science et Musique" (IRISA, Rennes, October 2019).

10.1.2. Scientific Events: Selection

10.1.2.1. Chair of Conference Program Committees

- Anatole Lécuyer was Member of the organization committee of IEEE VR 2019 (Panels).
- Ferran Argelaguet was Program Chair of the IEEE VR 2019 conference track.

10.1.2.2. Member of the Conference Program Committees

- Ferran Argelaguet was Member of the conference program committee of ACM SUI 2019.
- Valerie Gouranton was Member of the conference program committee of EuroVR 2019.
- Guillaume Moreau was Member of IEEE ISMAR 2019 Program committee.
- Jean-Marie Normand was Member of the program committee of IEEE VR 2019 Conference Track and of the IEEE AIVR 2019 (International Conference on Artificial Intelligence and Virtual Reality).

10.1.2.3. Reviewer

- Anatole Lécuyer was Reviewer for IEEE VRST 2019, IEEE VR 2019.
- Ferran Argelaguet was Reviewer for IEEE VR 2019, IEEE VRST 2019, ACM CHI 2019, ACM UIST 2019, IEEE ISMAR 2019.
- Maud Marchal was Reviewer for Eurographics 2019, ACM Siggraph Asia 2019, ACM UIST 2019, Eurohaptics 2019, IEEE Haptic Symposium 2019, IEEE ICRA 2019, IHM 2019.
- Valérie Gouranton was Reviewer for IEEE VR 2019, Euro VR 2019
- Guillaume Moreau was Reviewer for IEEE VR 2019, IEEE ISMAR 2019, ACM Augmented Human 2019 and IAPR Machine Vision Applications.
- Jean-Marie Normand was Reviewer for IEEE AIVR 2019, AH 2019, IEEE ISMAR 2019, IEEE VR 2019 Journal Track, IEEE VR 2019 Conference Track, IEEE CoG 2019 (Conference on Games), ACM SAP 2019.

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, Frontiers in Virtual Environments, and Presence journals.
- Valérie Gouranton is Review Editor of Frontiers in Virtual Environments.

- Ferran Argelaguet is Review Editor of Frontiers in Virtual Environments.
- Maud Marchal is Review Editor of Frontiers in Virtual Environments.
- Guillaume Moreau is Review Editor of Frontiers in Virtual Environments.
- Jean-Marie Normand is Review Editor of Frontiers in Virtual Environments.

10.1.3.2. Reviewer - Reviewing Activities

- Ferran Argelaguet was Reviewer for IEEE Transactions on Visualization and Computer Graphics, IEEE Computer Graphics and Applications, International Journal on Human Computer Studies.
- 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 Computers & Graphics, IEEE Transactions on Visualization and Computer Graphics, Frontiers in Virtual Environments.
- Jean-Marie Normand was Reviewer for IEEE Transactions on Visualization and Computer Graphics, Computer Animation and Virtual Worlds.
- Ronan Gaugne was Reviewer for Thermal Science and Engineering Progress, Elsevier and Archaeological and Anthropological Sciences, Springer

10.1.4. Invited Talks

- Valérie Gouranton was invited to give a talk at Aristote Association, Ecole Polytechnique Paris (June 2019) and at Consortium 3D of TGIR HumaNum, Nantes (Dec. 2019)
- Guillaume Moreau was invited to give invited talks at University of South Australia, at the Research workshop of Ecole Centrale Group and at IRT b<>com.
- Ronan Gaugne was invited to give a talk at the EVA (Electronic information, the Visual Arts and beyond) conference “L’original et la copie”, at Musée du Quai Branly, Paris, in October 2019.

10.1.5. Leadership within the Scientific Community

- Anatole Lécuyer is Member of the Scientific Board of INCR ("Institut des Neurosciences Cliniques de Rennes")
- Ronan Gaugne is Member of the Selection and Validation Committee for the French cluster “Pôle Images et Réseaux”, and of the Consortium 3D of TGIR HumaNum.
- Valérie Gouranton is Member of the Executive Committee of AFRV (French Association for Virtual Reality), and of the Consortium 3D of TGIR HumaNum.
- Maud Marchal is Member of the Executive Committee of Eurographics French Chapter.
- Guillaume Moreau is Member of the Steering Committee of IEEE ISMAR Conference.

10.1.6. Scientific Expertise

- Ferran Argelaguet was Member of a selection committee for the ANR.
- Guillaume Moreau is Member of ANSES (National Health Agency) Working Group on the “sanitary effects of Virtual, Mixed and Augmented Reality”. He is also Member of the HCERES (Higher Education and Research Evaluation Council) Committee of SIGMA Engineering School.

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 UMR 6074 and she is a member of the Conseil National des Universités 27th section (computer science).

- Guillaume Moreau is Dean of Studies at ECN.
- Jean-Marie Normand is Head of the “Virtual Reality” major at ECN.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Anatole Lécuyer:

Master AI: “Haptic Interaction and Brain-Computer Interfaces”, 6h, M2, Ecole Polytechnique, FR

Master MNRV: “Haptic Interaction”, 9h, M2, ENSAM, Laval, 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 SIF: “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”, 12h, 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: “Techniques d’Interaction Avancées”, 26h, M2, ISTIC, University of Rennes 1, FR

Master SIF: “Virtual Reality and Multi-Sensory Interaction”, 8h, M2, University of Rennes 1, FR

Master SIF: “Data Mining and Visualization”, 2h, M2, University of Rennes 1, 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

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”, 3h, M1, University of Rennes 2, FR

Ronan Gaugne:

INSA Rennes: Projects on “Virtual Reality”, 24h, L3, Insa Rennes, FR

Master Digital Creation: “Virtual Reality”, 6h, M1, University of Rennes 2, FR

Guillaume Moreau:

Dean of studies, Ecole Centrale de Nantes, FR

Virtual Reality Major, “C++ Programming for VR”, 30h, M1/M2, Ecole Centrale de Nantes, FR

Virtual Reality Major, “Fundamentals of Virtual Reality”, 6h, M1/M2, Ecole Centrale de Nantes, FR

Virtual Reality Major, “Computer Graphics”, 4h, M1/M2, Ecole Centrale de Nantes, FR

Virtual Reality Major, “Advanced Software Development”, 20h, M1/M2, Ecole Centrale de Nantes, FR

Computer Science Major, “Discrete Mathematics”, 10h, M1/M2, Ecole Centrale de Nantes, FR

Jean-Marie Normand:

Virtual Reality Major, “Computer Graphics”, 24h, M1/M2, Ecole Centrale de Nantes, FR

Virtual Reality Major, “Fundamentals of Virtual Reality”, 14h, M1/M2, Ecole Centrale de Nantes, FR

Virtual Reality Major, “Computer Vision and Augmented Reality”, 26h, M1/M2, Ecole Centrale de Nantes, FR

Virtual Reality Major, “Advanced Concepts”, 24h, M1/M2, Ecole Centrale de Nantes, FR

Virtual Reality Major, “Projects on Virtual Reality”, 20h, M1/M2, Ecole Centrale de Nantes, FR

10.2.2. Supervision

- PhD: Hakim Si-Mohammed, “Design and Study of Interactive Systems based on Brain Computer Interfaces and Augmented Reality”, INSA de Rennes, Defended December 3rd, 2019, Supervised by Anatole Lécuyer, Géry Casiez (Mjolnir, Inria) and Ferran Argelaguet
- PhD in progress: Hadrien Gurnel, “Assistance robotisée d’insertion d’aiguille par comanipulation”, Started in October 2016, Supervised by Alexandre Krupa (Rainbow, Inria) and Maud Marchal
- PhD in progress: 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)
- PhD in progress: Xavier de Tinguy, “Haptic manipulation in virtual environments”, Started in September 2017, Supervised by Maud Marchal, Claudio Pacchierotti (Rainbow, Inria) and Anatole Lécuyer
- PhD in progress: Rebecca Fribourg, “Perception and interaction with and via avatars”, Started in September 2017, Supervised by Ferran Argelaguet, Ludovic Hoyet (Mimetic, Inria) and Anatole Lécuyer
- PhD in progress: Romain Lagneau, “Data-driven models for dexterous manipulation of robots”, Started in Septembre 2017, Supervised by Maud Marchal and Alexandre Krupa (Rainbow, Inria)
- PhD in progress: Flavien Lécuyer, “Interactive digital introspection methods for archeology”, Started in September 2017, Supervised by Valérie Gouranton, Grégor Marchand (CNRS UMR CREAAH) and Bruno Arnaldi
- PhD in progress: Guillaume Bataille, “Natural interactions with IoT using VR/AR”, Started in October 2017, Supervised by Valérie Gouranton, Danielle Pelé and Jérémy Lacoche (Orange Labs) and Bruno Arnaldi
- PhD in progress: 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
- PhD in progress: Romain Terrier, “Presence of self and others in a collaborative virtual environment”, Started in October 2017, Supervised by Valérie Gouranton, Nico Pallamin (b<>com), Cédric Bach (HDG) and Bruno Arnaldi

- PhD in progress: Mathis Fleury, “Neurofeedback based on fMRI and EEG”, Started in November 2017, Supervised by Anatole Lécuyer and Christian Barillot (Visages, Inria)
- PhD in progress: Tiffany Luong, “Affective VR: acquisition, modelling, and exploitation of affective states in virtual reality”, Started in February 2018, Supervised by Anatole Lécuyer, Marc Diverrez (b<>com), Ferran Argelaguet
- PhD in progress: Hugo Brument, “Towards user-adapted interaction techniques based on human locomotion laws for navigating in virtual environments”, Started in October 2018, Supervised by Ferran Argelaguet, Maud Marchal and Anne-Hélène Olivier (MimeTIC, Inria)
- PhD in progress: Diane Dewez, “Avatar-Based Interaction in Virtual Reality”, Started in October 2018, Supervised by Anatole Lécuyer, Ferran Argelaguet and Ludovic Hoyet (MimeTIC)
- PhD in progress: Victor Rodrigo Mercado Garcia, “Encountered-type haptics”, Started in October 2018, Supervised by Maud Marchal and Anatole Lécuyer
- PhD in progress: Guillaume Vailland, “Outdoor wheelchair assisted navigation: reality versus virtuality”, Started in November 2018, Supervised by Valérie Gouranton and Marie Babel (Rainbow, Inria)
- PhD in progress: Gwendal Fouché, “Immersive Interaction and Visualization of Temporal 3D Data”, Started in October 2019, Supervised by Ferran Argelaguet, Charles Kervrann (Serpico Team) and Emmanuelle Faure (Mosaic Team).
- PhD in progress: Adelaide Genay, “Embodiment in Augmented Reality”, Started in October 2019, Supervised by Anatole Lécuyer, Martin Hachet (Potioc, Inria)
- PhD in progress: Martin Guy, “Physiological markers for characterizing virtual embodiment”, Started in October 2019, Supervised by Guillaume Moreau, Jean-Marie Normand and Camille Jeunet (CNRS, CLEE)
- PhD in progress: Grégoire Richard, “Touching Avatars: The role of haptic feedback in virtual embodiment”, Started in October 2019, Supervised by Géry Casiez (Loki, Inria), Thomas Pietzrak (Loki, Inria), Anatole Lécuyer and Ferran Argelaguet
- PhD in progress: Sebastian Vizcay, “Dexterous Interaction in Virtual Reality using High-Density Electrotactile Feedback”, Started in November 2019, Supervised by Ferran Argelaguet, Maud Marchal and Claudio Pacchierotti (Rainbow, Inria)

10.2.3. Juries

- Anatole Lécuyer was Referee for the PhD Theses of Justine Saint-Aubert (ISIR-UPMC), Geoffrey Gorisse (ENSAM), Léa Pillette (Univ. Bordeaux), Examiner for the PhD Thesis of Mélodie Fouillen (Inserm Lyon), Grégoire Cattan (GIPSA-Lab), and Examiner for the HDR Thesis of Sinan Haliyo (ISIR-UPMC).
- Bruno Arnaldi was Referee for the PhD Theses of Rémi Lacaze-Labadie (UTC Compiègne) and Nicolas Muller (IRIT, Toulouse), external reviewer for the PhD defense of Alexandre Kabil (IMT Atlantique - Brest) and member for the PhD Defense of Hakim Si Mohammed (INSA Rennes).
- Guillaume Moreau was external reviewer for the PhD defense of Nicolas Muller (IRIT, Univ. Toulouse), Alice Guerville-Ballé (Univ. Pau), Jason Rambach (Univ. Kaiserslautern, DE), president of the committee for Thibaud Toullier (IFSTTAR, Univ. Rennes), Nam Duong Duong (IRT b<>com, Centrale-Supelec).
- Jean-Marie Normand was examiner for the PhD Thesis of Geoffrey Gorisse (ENSAM).

10.3. Popularization

10.3.1. Articles and contents

- Ouest-France: Interview of Anatole Lécuyer (October 2019)

- France Inter: Interview of Anatole Lécuyer in the broadcast "Du vent dans les synapses" (November 2019)
- France TV: Interview of Guillaume Moreau and Jean-Marie Normand in "Télématin"

10.3.2. Interventions

- “Technoférence Humain Augmenté”: Presentation from Rebecca Fribourg (Nantes, January 2019)
- “Made by DV-Group”: Event organized by Rennes Metropole about “VR and Cinema”, presentations and demos from Ferran Argelaguet (Rennes, February 2019)
- “Laval Virtual 2019”: Demos of FIVE/SEVEN software (March 2019)
- “Journées Nationales de l’Archéologie”: Demos made by Ronan Gaugne and Valérie Gouranton (Champs Libres, Rennes, June 2019)
- “Journées Européennes du Patrimoine”: Conference and Demos made by Ronan Gaugne and Valérie Gouranton (Musée des Beaux Arts, Rennes, September 2019)
- “Festival Demain/Maintenant” : Presentation from Anatole Lécuyer (Rennes, October 2019)
- “Fête de la Science” 2019 : Presentation from Anatole Lécuyer (Saint-Louis, October 2019)
- “Journée Science et Musique” : Demos made by Florian Nouviale, Ronan Gaugne and Valérie Gouranton (Rennes, October 2019). Rebecca Fribourg and Diane Dewez also participated in the organization of the event.
- “Festival du Livre du Var” 2019 : Presentation from Anatole Lécuyer (Toulon, November 2019)
- “Digital Design Days” 2019 : Presentation from Ronan Gaugne (Milano, Italy, November 2019)

10.3.3. Internal action

- Inria/IRISA “VR Hackathon” 2019: Hackathon (20 participants) organized by Hybrid team and open to all members of Inria Rennes center (May 2019).
- “20 years of Immersia”: Event organized by Inria Rennes/IRISA center, with presentations from Anatole Lécuyer and Bruno Arnaldi, and demos made by Florian Nouviale and Ronan Gaugne (November 2019).

10.3.4. Creation of media or tools for science outreach

The ANR-FRQSC INTROSPECT project produced a transparent 3D printing of the content of an Egyptian mummy cat from a CT scan. The mummy is part of the collection of the Musée des Beaux-Arts of Rennes. The 3D printing is permanently exhibited in the Museum, next to the real mummy (Figure 21)

The immersive VR application of the LSI project developped during the hosting of the invited professor Franz Fischnaller from Albertina Accademy of Fine Arts around the Last Supper painting of Lenonardo da Vinci was deployed and exhbited in the Deep Space 8K of the Ars Electronica center in Linz, Austria (Figure 22).

11. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] J.-M. Batail, S. Bioulac, F. Cabestaing, C. Daudet, D. Drapier, M. Fouillen, T. Fovet, A. Hakoun, R. Jardri, C. Jeunet, F. Lotte, E. Maby, J. Mattout, T. Medani, J.-A. Micoulaud-Franchi, J. Mladenović, L. Perronet, L. Pillette, T. Ros, F. Vialatte. *EEG neurofeedback research A fertile ground for psychiatry?*, in "L’Encéphale", 2019, vol. 45, n^{o} 3, p. 245-255 [*DOI :* 10.1016/j.encep.2019.02.001], https://hal-univ-rennes1.archives-ouvertes.fr/hal-02094863

Figure 21. The mummy cat and its copy in the Museum des Beaux-Arts in Rennes.

Figure 22. The LSI project in the Ars Electronica center in Linz.

[2] M.-S. BRACQ, E. MICHINOV, B. ARNALDI, B. CAILLAUD, B. GIBAUD, V. GOURANTON, P. JANNIN.*Learning procedural skills with a virtual reality simulator An acceptability study*, in "Nurse Education Today", August 2019, vol. 79, p. 153-160 [*DOI :* 10.1016/J.NEDT.2019.05.026], https://hal-univ-rennes1.archives-ouvertes.fr/hal-02150192

[3] A. COSTES, F. ARGELAGUET SANZ, F. DANIEAU, P. GUILLOTEL, A. LÉCUYER.*Touchy : A Visual Approach for Simulating Haptic Effects on Touchscreens*, in "Frontiers in information and communication technologies", February 2019, vol. 6, p. 1-11 [*DOI :* 10.3389/FICT.2019.00001], https://hal.inria.fr/hal-02408724

[4] Y. GAO, Y. LIU, J.-M. NORMAND, G. MOREAU, X. GAO, Y. WANG.*A study on differences in human perception between a real and an AR scene viewed in an OST-HMD*, in "Journal of the Society for Information Display", January 2019, p. 1-17 [*DOI :* 10.1002/JSID.752], https://hal.inria.fr/hal-01987255

[5] Y. GAO, E. PEILLARD, J.-M. NORMAND, G. MOREAU, Y. LIU, Y. WANG.*Influence of virtual objects' shadows and lighting coherence on distance perception in optical see-through augmented reality*, in "Journal of the Society for Information Display", July 2019 [*DOI :* 10.1002/JSID.832], https://hal.archives-ouvertes.fr/hal-02200576

[6] T. HOWARD, M. MARCHAL, A. LÉCUYER, C. PACCHIEROTTI.*PUMAH : Pan-tilt Ultrasound Mid-Air Haptics for larger interaction workspace in virtual reality*, in "IEEE Transactions on Haptics (ToH)", January 2020, p. 1-6 [*DOI :* 10.1109/TOH.2019.2963028], https://hal.inria.fr/hal-02424247

[7] N. KOSMYNA, A. LÉCUYER.*A conceptual space for EEG-based brain- computer interfaces*, in "PLoS ONE", January 2019, p. 1-30 [*DOI :* 10.1371/JOURNAL.PONE.0210145], https://hal.inria.fr/hal-02394251

[8] M. LE CHÉNÉCHAL, T. DUVAL, V. GOURANTON, J. ROYAN, B. ARNALDI.*Help! I Need a Remote Guide in my Mixed Reality Collaborative Environment*, in "Frontiers in Robotics and AI", 2019, p. 1-25, forthcoming [*DOI :* 10.3389/FROBT.2019.00106], https://hal.archives-ouvertes.fr/hal-02314193

[9] C. LOMBART, E. MILLAN, J.-M. NORMAND, A. VERHULST, B. LABBÉ-PINLON, G. MOREAU.*Consumer perceptions and purchase behavior toward imperfect fruits and vegetables in an immersive virtual reality grocery store*, in "Journal of Retailing and Consumer Services", May 2019, vol. 48, p. 28-40 [*DOI :* 10.1016/J.JRETCONSER.2019.01.010], https://hal.archives-ouvertes.fr/hal-01995916

International Conferences with Proceedings

[10] N. BAZOGE, R. GAUGNE, F. NOUVIALE, V. GOURANTON, B. BOSSIS.*Expressive potentials of motion capture in the Vis Insita musical performance*, in "NIME 2019 - The International Conference on New Interfaces for Musical Expression", Porto Alegre, Brazil, June 2019, p. 1-5, https://hal.archives-ouvertes.fr/hal-02297397

[11] H. BRUMENT, R. FRIBOURG, G. GALLAGHER, T. M. HOWARD, F. LÉCUYER, T. LUONG, V. MERCADO, E. PEILLARD, X. DE TINGUY, M. MARCHAL.*Pyramid Escape: Design of Novel Passive Haptics Interactions for an Immersive and Modular Scenario*, in "VR 2019 - 26th IEEE Conference on Virtual Reality and 3D User Interfaces", Osaka, Japan, IEEE, March 2019, p. 1409-1410 [*DOI :* 10.1109/VR.2019.8797848], https://hal.inria.fr/hal-02284129

[12] H. BRUMENT, I. PODKOSOVA, H. KAUFMANN, A.-H. OLIVIER, F. ARGELAGUET SANZ.*Virtual vs. Physical Navigation in VR: Study of Gaze and Body Segments Temporal Reorientation Behaviour*, in "VR

2019 - 26th IEEE Conference on Virtual Reality and 3D User Interfaces", Osaka, Japan, IEEE, March 2019, p. 680-689 [*DOI :* 10.1109/VR.2019.8797721], https://hal.inria.fr/hal-02058365

[13] S. BUTET, G. LIOI, M. FLEURY, A. LÉCUYER, C. BARILLOT, I. BONAN.*A multi-target motor imagery training using EEG-fMRI Neurofeedback: an exploratory study on stroke*, in "OHBM 2019- Orgaization for Human Brain Mapping", Rome, Italy, June 2019, p. 1-4, https://hal.inria.fr/hal-02265496

[14] X. DE TINGUY, C. PACCHIEROTTI, M. EMILY, M. CHEVALIER, A. GUIGNARDAT, M. GUILLAUDEUX, C. SIX, A. LÉCUYER, M. MARCHAL.*How different tangible and virtual objects can be while still feeling the same?*, in "WHC 2019 - IEEE World Haptics Conference", Tokyo, Japan, IEEE, July 2019, p. 1-6 [*DOI :* 10.1109/WHC.2019.8816164], https://hal.inria.fr/hal-02121839

[15] X. DE TINGUY, C. PACCHIEROTTI, M. MARCHAL, A. LÉCUYER.*Toward Universal Tangible Objects: Optimizing Haptic Pinching Sensations in 3D Interaction*, in "VR 2019 - 26th IEEE Conference on Virtual Reality and 3D User Interfaces", Osaka, Japan, IEEE, 2019, p. 321-330 [*DOI :* 10.1109/VR.2019.8798205], https://hal.inria.fr/hal-02021319

[16] D. DEWEZ, R. FRIBOURG, F. ARGELAGUET SANZ, L. HOYET, D. MESTRE, M. SLATER, A. LÉCUYER.*Influence of Personality Traits and Body Awareness on the Sense of Embodiment in Virtual Reality*, in "ISMAR 2019 - 18th IEEE International Symposium on Mixed and Augmented Reality", Beijin, China, October 2019, p. 1-12, https://hal.inria.fr/hal-02385783

[17] Q. GALVANE, I.-S. LIN, F. ARGELAGUET SANZ, T.-Y. LI, M. CHRISTIE.*VR as a Content Creation Tool for Movie Previsualisation*, in "VR 2019 - 26th IEEE Conference on Virtual Reality and 3D User Interfaces", Osaka, France, IEEE, March 2019, p. 303-311 [*DOI :* 10.1109/VR.2019.8798181], https://hal.inria.fr/hal-02295309

[18] *Best Paper*
R. GAUGNE, T. NICOLAS, Q. PETIT, M. OTSUKI, V. GOURANTON.*Evaluation of a Mixed Reality based Method for Archaeological Excavation Support*, in "ICAT-EGVE 2019 - International Conference on Artificial Reality and Telexistence - Eurographics Symposium on Virtual Environments", Tokyo, Japan, September 2019, p. 1-8, https://hal.inria.fr/hal-02272910.

[19] R. GAUGNE, Q. PETIT, J.-B. BARREAU, V. GOURANTON.*Interactive and Immersive Tools for Point Clouds in Archaeology*, in "ICAT-EGVE 2019 - International Conference on Artificial Reality and Telexistence - Eurographics Symposium on Virtual Environments", Tokyo, Japan, September 2019, p. 1-8, https://hal.inria.fr/hal-02272912

[20] H. GURNEL, M. MARCHAL, L. LAUNAY, L. BEUZIT, A. KRUPA.*Design of haptic guides for pre-positioning assistance of a comanipulated needle*, in "SMC 2019 - IEEE International Conference on Systems, Man, and Cybernetics", Bari, Italy, IEEE, October 2019, p. 478-485 [*DOI :* 10.1109/SMC.2019.8914395], https://hal.inria.fr/hal-02387192

[21] H. GURNEL, M. MARCHAL, L. LAUNAY, L. BEUZIT, A. KRUPA.*Preliminary evaluation of haptic guidance for pre-positioning a comanipulated needle*, in "SURGETICA 2019", Rennes, France, June 2019, p. 1-3, https://hal.inria.fr/hal-02387214

[22] T. HOWARD, G. GALLAGHER, A. LÉCUYER, C. PACCHIEROTTI, M. MARCHAL.*Investigating the recognition of local shapes using mid-air ultrasound haptics*, in "WHC 2019 - IEEE World Haptics Conference", Tokyo, Japan, IEEE, July 2019, p. 1-6 [*DOI :* 10.1109/WHC.2019.8816127], https://hal.inria.fr/hal-02121329

[23] *Best Paper*
J. LACOCHE, T. DUVAL, B. ARNALDI, E. MAISEL, J. ROYAN.*Machine Learning Based Interaction Technique Selection For 3D User Interfaces*, in "EuroVR 2019 - 16th EuroVR International Conference", Tallinn, Estonia, Springer, October 2019, p. 33-51 [*DOI :* 10.1007/978-3-030-31908-3_3], https://hal.archives-ouvertes.fr/hal-02292434.

[24] J. LACOCHE, T. DUVAL, B. ARNALDI, E. MAISEL, J. ROYAN.*3DPlasticToolkit: Plasticity for 3D User Interfaces*, in "EuroVR 2019 - 16th EuroVR International Conference", Tallinn, Estonia, Springer, October 2019, vol. 11883, p. 62-83 [*DOI :* 10.1007/978-3-030-31908-3_5], https://hal.archives-ouvertes.fr/hal-02292436

[25] G. LIOI, S. BUTET, M. FLEURY, C. CURY, E. BANNIER, A. LÉCUYER, I. BONAN, C. BARILLOT.*Bimodal EEG-fMRI Neurofeedback for upper motor limb rehabilitation: a pilot study on chronic patients*, in "rtFIN 2019 - Real Time Functional Imaging and Neurofeedback", Maastricht, Netherlands, December 2019, p. 1-2, https://hal.inria.fr/hal-02383532

[26] T. LUONG, N. MARTIN, F. ARGELAGUET SANZ, A. LÉCUYER.*Studying the Mental Effort in Virtual Versus Real Environments*, in "VR 2019 - 26th IEEE Conference on Virtual Reality and 3D User Interfaces", Osaka, France, IEEE, March 2019, p. 809-816 [*DOI :* 10.1109/VR.2019.8798029], https://hal.inria.fr/hal-02295331

[27] F. LÉCUYER, V. GOURANTON, A. REUZEAU, R. GAUGNE, B. ARNALDI.*Authoring AR Interaction by AR*, in "ICAT-EGVE 2019 - International Conference on Artificial Reality and Telexistence - Eurographics Symposium on Virtual Environments", Tokyo, Japan, September 2019, p. 1-8, https://hal.inria.fr/hal-02272930

[28] F. LÉCUYER, V. GOURANTON, A. REUZEAU, R. GAUGNE, B. ARNALDI.*Create by doing - Action sequencing in VR*, in "CGI 2019 - Computer Graphics International", Calgary, Canada, Springer, 2019, p. 329-335 [*DOI :* 10.1007/978-3-030-22514-8_27], https://hal.archives-ouvertes.fr/hal-02119288

[29] M.-A. PARADIS, T. NICOLAS, R. GAUGNE, J.-B. BARREAU, R. AUGER, V. GOURANTON.*Making virtual archeology great again (without scientific compromise)*, in "CIPA 2019 - 27th CIPA International Symposium - Documenting the Past for a Better Future", Ávila, Spain, D. GONZALEZ-AGUILERA, F. REMONDINO, I. TOSCHI, P. RODRIGUEZ-GONZALVEZ, E. STATHOPOULOU (editors), International Society for Photogrammetry and Remote Sensing, 2019, vol. 42, n^{o} 2/W15, p. 879-886 [*DOI :* 10.5194/ISPRS-ARCHIVES-XLII-2-W15-879-2019], https://hal-univ-rennes1.archives-ouvertes.fr/hal-02310425

[30] E. PEILLARD, F. ARGELAGUET SANZ, J.-M. NORMAND, A. LÉCUYER, G. MOREAU.*Studying Exocentric Distance Perception in Optical See-Through Augmented Reality*, in "ISMAR 2019 - 18th IEEE International Symposium on Mixed and Augmented Reality", Beijing, China, October 2019, p. 1-8, https://hal.archives-ouvertes.fr/hal-02200822

[31] *Best Paper*
E. PEILLARD, T. THEBAUD, J.-M. NORMAND, F. ARGELAGUET SANZ, G. MOREAU, A. LÉCUYER.*Virtual Objects Look Farther on the Sides: The Anisotropy of Distance Perception in Virtual Reality*, in "VR 2019 - 26th IEEE Conference on Virtual Reality and 3D User Interfaces", Osaka, Japan, IEEE, March 2019, p. 227-236 [*DOI :* 10.1109/VR.2019.8797826], https://hal.archives-ouvertes.fr/hal-02084069.

[32] H. SI-MOHAMMED, G. CASIEZ, F. ARGELAGUET SANZ, N. ROUSSEL, A. LÉCUYER.*Defining Brain-Computer Interfaces: A Human-Computer Interaction Perspective*, in "GBCIC 2019 - 8th Graz Brain-Computer Interface Conference", Graz, Austria, September 2019, p. 1-4, https://hal.inria.fr/hal-02163771

[33] R. TERRIER, N. MARTIN, J. LACOCHE, V. GOURANTON, B. ARNALDI.*Am I better in VR with a real audience?*, in "CGI 2019 - Computer Graphics International", Calgary, Canada, Springer, June 2019, p. 28-39 [*DOI :* 10.1007/978-3-030-22514-8_3], https://hal.archives-ouvertes.fr/hal-02119308

[34] G. VAILLAND, F. GRZESKOWIAK, L. DEVIGNE, Y. GAFFARY, B. FRAUDET, E. LEBLONG, F. NOUVIALE, F. PASTEAU, R. LE BRETON, S. GUEGAN, V. GOURANTON, B. ARNALDI, M. BABEL.*User-centered design of a multisensory power wheelchair simulator: towards training and rehabilitation applications*, in "ICORR 2019 - International Conference on Rehabilitation Robotics", Toronto, Canada, 2019, p. 1-6, https://hal.inria.fr/hal-02134530

National Conferences with Proceeding

[35] R. GAUGNE, V. GOURANTON.*3D pour l'archéologie : quelles interactions, pour quoi faire ?*, in "2019 - Archéologie : imagerie numérique et 3D", Rennes, France, INRAP (editor), Séminaire scientifique et technique de l'Inrap, Sylvie Eusèbe and Théophane Nicolas and Valérie Gouranton and Ronan Gaugne, May 2019, vol. 3, https://sstinrap.hypotheses.org/1429 [*DOI :* 10.34692/XVYT-HR49], https://hal-inrap.archives-ouvertes.fr/hal-02121567

Conferences without Proceedings

[36] A. BERNARDIN, C. DURIEZ, M. MARCHAL.*An Interactive Physically-based Model for Active Suction Phenomenon Simulation*, in "SWS19 - SOFA Week Symposium", Paris, France, November 2019, https://hal.inria.fr/hal-02419381

[37] L. DEVIGNE, M. BABEL, R. LE BRETON, E. LEBLONG, V. GOURANTON, F. PASTEAU, B. FRAUDET, S. GUEGAN, G. VAILLAND, Y. GAFFARY.*Expérience de conduite en fauteuil roulant dans une ville virtuelle avec un simulateur multisensoriel conçu selon une approche centrée sur l'utilisateur : une étude de cas*, in "SOFMER 2019 - 34ème congrès de la Société Français de Médecine Physique et de Réadaptation", Bordeaux, France, October 2019, https://hal.inria.fr/hal-02339573

[38] M. FLEURY, G. LIOI, C. BARILLOT, A. LÉCUYER.*The use of haptic feedback in Brain-Computer Interfaces and Neurofeedback*, in "CORTICO 2019 - COllectif pour la Recherche Transdisciplinaire sur les Interfaces Cerveau-Ordinateur", Lille, France, March 2019, https://hal.archives-ouvertes.fr/hal-02387408

[39] M. FLEURY, G. LIOI, C. BARILLOT, A. LÉCUYER.*The use of haptic feedback in Brain-Computer Interfaces and Neurofeedback*, in "rtFIN 2019 - Real Time Functional Imaging and Neurofeedback", Maastricht, Netherlands, December 2019, https://hal.archives-ouvertes.fr/hal-02387400

[40] S. LE FRANC, M. FLEURY, M. COGNÉ, S. BUTET, C. BARILLOT, A. LÉCUYER, I. BONAN.*Influence of visual feedback on the illusion of movement induced by tendon vibration of wrist in healthy subjects*, in "SOFMER 2019 - 34ème congrès de la Société Français de Médecine Physique et de Réadaptation", Bordeaux, France, October 2019, https://hal.inria.fr/hal-02415992

[41] G. LIOI, S. BUTET, M. FLEURY, A. LÉCUYER, I. BONAN, C. BARILLOT.*Efficacy of EEG-fMRI Neurofeedback in stroke in relation to the DTI structural damage: a pilot study*, in "OHBM 2019 - 25th Annual Meeting of the Organization for Human Brain Mapping", Rome, Italy, June 2019, p. 1-4, https://hal.inria.fr/hal-02265495

Scientific Books (or Scientific Book chapters)

[42] J.-B. BARREAU, E. LANOË, R. GAUGNE.*3D Sketching of the Fortified Entrance of the Citadel of Aleppo from a Few Sightseeing Photos*, in "Digital Cultural Heritage", H. KREMERS (editor), Springer International Publishing, June 2020, p. 359-371, First Online : 22 June 2019 [*DOI :* 10.1007/978-3-030-15200-0_24], https://hal-univ-rennes1.archives-ouvertes.fr/hal-02164770

Other Publications

[43] M.-S. BRACQ, E. MICHINOV, B. ARNALDI, A. AUDINOT, B. CAILLAUD, B. GIBAUD, V. GOURANTON, P.-L. HENAUX, A. LAMERCERIE, B. NOGUES, P. JANNIN.*Modeling, Simulation and Training Procedural Skills: User experience and acceptability of a virtual reality simulator for scrub nurses in neurosurgery*, January 2019, vol. 36, 1, 19th International Meeting on Simulation in Healthcare (IMSH), Poster, https://hal.archives-ouvertes.fr/hal-02123682

References in notes

[44] D. A. BOWMAN, E. KRUIJFF, J. J. LAVIOLA, I. POUPYREV.*3D User Interfaces: Theory and Practice*, Addison Wesley, 2004

[45] 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^{o} 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

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
A2.1.9. - 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. - Formal method for 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. - Methods in 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. Team, Visitors, External Collaborators

Research Scientists
Benoît Caillaud [Team leader, Inria, Senior Researcher, HDR]
Albert Benveniste [Inria, Emeritus, HDR]
Khalil Ghorbal [Inria, Researcher]

Technical Staff
Mathias Malandain [Inria, Engineer]

PhD Students
Christelle Kozaily [Inria, PhD Student]
Aurélien Lamercerie [Univ de Rennes I, PhD Student, also with SemLis team of IRISA]
Joan Thibault [Univ de Rennes I, PhD Student, from Sep 2019]

Post-Doctoral Fellow
Benoît Vernay [Inria, Post-Doctoral Fellow, until Mar 2019]

Administrative Assistant
Armelle Mozziconacci [CNRS, Administrative Assistant]

2. Overall Objectives

2.1. Overall Objectives

Hycomes was created a local team of the Rennes — Bretagne Atlantique Inria research center in 2013 and has been created as an Inria Project-Team in 2016. The team is focused on two topics in cyber-physical systems design:

- 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. The objective of our research is to bridge the gap between system-level requirements, often expressed in natural, constrained or semi-formal languages and formal models, that can be simulated and verified.

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 [0]. 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 [0]. A wider set of tools, both industrial and academic, now exists in this segment [0]. In the EDA sector, VHDL-AMS was developed as a standard [12] and also allows for differential algebraic equations. Several domain-specific languages and tools for mechanical systems or electronic circuits also support some restricted classes of differential algebraic equations. Spice is the historic and most striking instance of these domain-specific languages/tools [0]. The main difference is that equations are hidden and the fixed structure of the differential algebraic results from the physical domain covered by these languages.

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 [19], [15] 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 [15], [19], [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 [2], 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 [49].

[0] http://www.lccc.lth.se/media/LCCC2012/WorkshopSeptember/slides/Astrom.pdf
[0] https://www.modelica.org/
[0] SimScape by Mathworks, Amesim by LMS International, now Siemens PLM, and more.
[0] http://bwrcs.eecs.berkeley.edu/Classes/IcBook/SPICE/MANUALS/spice3.html

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 ∂ 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 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 [58], [41], [11]. 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. [45] 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. Structural Analysis of DAE Systems

The Modelica language is based on Differential Algebraic Equations (DAE). The general form of a DAE is given by:

$$F(t, x, x', x'', \cdots) \tag{1}$$

where F is a system of n_e equations $\{f_1, \cdots, f_{n_e}\}$ and x is a finite list of n_v independent real-valued, smooth enough, functions $\{x_1, \cdots, x_{n_v}\}$ of the independent variable t. We use x' as a shorthand for the list of first-order time derivatives of x_j, $j = 1, \cdots, n_v$. High-order derivatives are recursively defined as usual, and $x^{(k)}$ denotes the list formed by the k-th derivatives of the functions x_j. Each f_i depends on the scalar t and some of the functions x_j as well as a finite number of their derivatives.

Let $\sigma_{i,j}$ denote the highest differentiation order of variable x_j effectively appearing in equation f_i, or $-\infty$ if x_j does not appear in f_i. The *leading variables* of F are the variables in the set

$$\left\{ x_j^{(\sigma_j)} \mid \sigma_j = \max_i \sigma_{i,j} \right\}$$

The *state variables* of F are the variables in the set

$$\left\{ x_j^{(\nu_j)} \mid 0 \le \nu_j < \max_i \sigma_{i,j} \right\}$$

A leading variable $x_j^{(\sigma_j)}$ is said to be *algebraic* if $\sigma_j = 0$ (in which case, neither x_j nor any of its derivatives are state variables). In the sequel, v and u denote the leading and state variables of F, respectively.

DAE are a strict generalization of *ordinary differential equations* (*ODE*), in the sense that it may not be immediate to rewrite a DAE as an explicit ODE of the form $v = G(u)$. The reason is that this transformation relies on the Implicit Function Theorem, requiring that the Jacobian matrix $\frac{\partial F}{\partial v}$ have full rank. This is, in general, not the case for a DAE. Simple examples, like the two-dimensional fixed-length pendulum in Cartesian coordinates [55], exhibit this behaviour.

For a square DAE of dimension n (i.e., we now assume $n_e = n_v = n$) to be solved in the neighborhood of some (v^*, u^*), one needs to find a set of non-negative integers $C = \{c_1, \cdots, c_n\}$ such that system

$$F^{(C)} = \{f_1^{(c_1)}, \cdots, f_n^{(c_n)}\}$$

can locally be made explicit, i.e., the Jacobian matrix of $F^{(C)}$ with respect to its leading variables, evaluated at (v^*, u^*), is nonsingular. The smallest possible value of $\max_i c_i$ for a set C that satisfies this property is the *differentiation index* [32] of F, that is, the minimal number of time differentiations of all or part of the equations f_i required to get an ODE.

In practice, the problem of automatically finding a "minimal" solution C to this problem quickly becomes intractable. Moreover, the differentiation index may depend on the value of (v^*, u^*). This is why, in lieu of numerical nonsingularity, one is interested in the *structural nonsingularity* of the Jacobian matrix, i.e., its almost certain nonsingularity when its nonzero entries vary over some neighborhood. In this framework, the *structural analysis* (*SA*) of a DAE returns, when successful, values of the c_i that are independent from a given value of (v^*, u^*).

A renowned method for the SA of DAE is the *Pantelides method*; however, Pryce's Σ*-method* is introduced also in what follows, as it is a crucial tool for our works.

3.3.1. Pantelides method

In 1988, Pantelides proposed what is probably the most well-known SA method for DAE [55]. The leading idea of his work is that the structural representation of a DAE can be condensed into a bipartite graph whose left nodes (resp. right nodes) represent the equations (resp. the variables), and in which an edge exists if and only if the variable occurs in the equation.

By detecting specific subsets of the nodes, called *Minimally Structurally Singular* (*MSS*) subsets, the Pantelides method iteratively differentiates part of the equations until a perfect matching between the equations and the leading variables is found. One can easily prove that this is a necessary and sufficient condition for the structural nonsingularity of the system.

The main reason why the Pantelides method is not used in our work is that it cannot efficiently be adapted to multimode DAE (mDAE). As a matter of fact, the adjacency graph of a mDAE has both its nodes and edges parametrized by the subset of modes in which they are active; this, in turn, requires that a parametrized Pantelides method must branch every time no mode-independent MSS is found, ultimately resulting, in the worst case, in the enumeration of modes.

3.3.2. Pryce's Σ-method

Albeit less renowned that the Pantelides method, Pryce's Σ-method [56] is an efficient SA method for DAE, whose equivalence to the Pantelides method has been proved by the author. This method consists in solving two successive problems, denoted by primal and dual, relying on the Σ*-matrix*, or *signature matrix*, of the DAE F.

This matrix is given by:

$$\Sigma = (\sigma_{ij})_{1 \leq i,j \leq n} \tag{2}$$

where σ_{ij} is equal to the greatest integer k such that $x_j^{(k)}$ appears in f_i, or $-\infty$ if variable x_j does not appear in f_i. It is the adjacency matrix of a weighted bipartite graph, with structure similar to the graph considered in the Pantelides method, but whose edges are weighted by the highest differentiation orders. The $-\infty$ entries denote non-existent edges.

The *primal problem* consists in finding a *maximum-weight perfect matching* (*MWPM*) in the weighted adjacency graph. This is actually an assignment problem, for the solving of which several standard algorithms exist, such as the push-relabel algorithm [44] or the Edmonds-Karp algorithm [43] to only give a few. However, none of these algorithms are easily parametrizable, even for applications to mDAE systems with a fixed number of variables.

The *dual problem* consists in finding the component-wise minimal solution $(C, D) = (\{c_1, \cdots, c_n\}, \{d_1, \cdots, d_n\})$ to a given linear programming problem, defined as the dual of the aforementioned assignment problem. This is performed by means of a *fixpoint iteration* (*FPI*) that makes use of the MWPM found as a solution to the primal problem, described by the set of tuples $\{(i, j_i)\}_{i \in \{1, \cdots, n\}}$:

1. Initialize $\{c_1, \cdots, c_n\}$ to the zero vector.
2. For every $j \in \{1, \cdots, n\}$,
$$d_j \longleftarrow \max_i (\sigma_{ij} + c_i)$$
3. For every $i \in \{1, \cdots, n\}$,
$$c_i \longleftarrow d_{j_i} - \sigma_{i,j_i}$$
4. Repeat Steps 2 and 3 until convergence is reached.

From the results proved by Pryce in [56], it is known that the above algorithm terminates if and only if it is provided a MWPM, and that the values it returns are independent of the choice of a MWPM whenever there exist several such matchings. In particular, a direct corollary is that the Σ-method succeeds as long as a perfect matching can be found between equations and variables.

Another important result is that, if the Pantelides method succeeds for a given DAE F, then the Σ-method also succeeds for F and the values it returns for C are exactly the differentiation indices for the equations that are returned by the Pantelides method. As for the values of the d_j, being given by $d_j = \max_i (\sigma_{ij} + c_i)$, they are the differentiation indices of the leading variables in $F^{(C)}$.

Working with this method is natural for our works, since the algorithm for solving the dual problem is easily parametrizable for dealing with multimode systems, as shown in our recent paper [31].

3.3.3. Block triangular decomposition

Once structural analysis has been performed, system $F^{(C)}$ can be regarded, for the needs of numerical solving, as an algebraic system with unknowns $x_j^{(d_j)}$, $j = 1 \cdots n$. As such, (inter)dependencies between its equations must be taken into account in order to put it into block triangular form (BTF). Three steps are required:

1. the *dependency graph* of system $F^{(C)}$ is generated, by taking into account the perfect matching between equations $f_i^{(c_i)}$ and unknowns $x_j^{(d_j)}$;
2. the *strongly connected components* (*SCC*) in this graph are determined: these will be the *equation blocks* that have to be solved;
3. the *block dependency graph* is constructed as the condensation of the dependency graph, from the knowledge of the SCC; a BTF of system $F^{(C)}$ can be made explicit from this graph.

3.4. 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 [10]. 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 [53]. 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 [3]. 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/G-contracts 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 [46], [35], [52], [14], [37]. Contract theories were initially developed as specification formalisms able to refuse some inputs from the environment [42]. A/G-contracts were advocated in [18] and are is still a very active research topic, with several contributions dealing with the timed [24] and probabilistic [29], [30] viewpoints in system design, and even mixed-analog circuit design [54].

Automata theoretic interfaces. Interfaces combine assumptions and guarantees in a single, automata theoretic specification. Most interface theories are based on Lynch Input/Output Automata [51], [50]. Interface Automata [61], [60], [62], [33] 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 [57] inherit from both Interface Automata and the originally unrelated notion of Modal Transition System [48], [13], [27], [47]. 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 [63], [21], [23], [39], [38], [22], probabilistic [29], [40] and energy-aware [34] 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 [59]. Most requirements engineering tools offer a poor structuring of the requirements and cannot be considered as formal modeling frameworks today. They are nothing less, but nothing more than an informal structured 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 were working on the development of the fly-by-wire and of the landing gear subsystems, leading to a long and cahotic convergence of the design process.

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. Highlights of the Year

4.1. Highlights of the Year

The Hycomes team has reached in 2019 an important milestone in the team's research objectives: the design and implementation of an implicit structural analysis algorithm supporting multimode DAE systems. This method is based on an encoding of the varying structure of a multimode DAE as Boolean functions, represented with Binary Decision Diagrams (BDD). This enables a complete strucutral analysis of a multimode DAE system, without enumerating its modes.

5. New Software and Platforms

5.1. 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

5.2. 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/

5.3. IsamDAE

Implicit Structural Analysis of Multimode DAE systems

KEYWORDS: Structural analysis - Differential algebraic equations - Multimode - Scheduling

SCIENTIFIC DESCRIPTION: Modeling languages and tools based on Differential Algebraic Equations (DAE) bring several specific issues that do not exist with modeling languages based on Ordinary Differential Equations. The main problem is the determination of the differentiation index and latent equations. Prior to generating simulation code and calling solvers, the compilation of a model requires a structural analysis step, which reduces the differentiation index to a level acceptable by numerical solvers.

The Modelica language, among others, allows hybrid models with multiple modes, mode-dependent dynamics and state-dependent mode switching. These Multimode DAE (mDAE) systems are much harder to deal with. The main difficulties are (i) the combinatorial explosion of the number of modes, and (ii) the correct handling of mode switchings.

The aim of the software is on the first issue, namely: How can one perform a structural analysis of an mDAE in all possible modes, without enumerating these modes? A structural analysis algorithm for mDAE systems has been designed and implemented, based on an implicit representation of the varying structure of an mDAE. It generalizes J. Pryce's Sigma-method to the multimode case and uses Binary Decision Diagrams (BDD) to represent the mode-dependent structure of an mDAE. The algorithm determines, as a function of the mode, the set of latent equations, the leading variables and the state vector. This is then used to compute a mode-dependent block-triangular decomposition of the system, that can be used to generate simulation code with a mode-dependent scheduling of the blocks of equations.

FUNCTIONAL DESCRIPTION: IsamDAE (Implicit Structural Analysis of Multimode DAE systems) is a software library for testing new structural analysis algorithms for multimode DAE systems, based on an implicit representation of incidence graphs, matchings between equations and variables, and block decompositions. The input of the software is a variable dimension multimode DAE system consisting in a set of guarded equations and guarded variable declarations. It computes a mode-dependent structural index-reduction of the multimode system and produces a mode-dependent graph for the scheduling of blocks of equations. Evaluation functions make it possible to return the lists of leading equations and leading variables, as well as the actual scheduling of blocks, in a specified mode.

IsamDAE is coded in OCaml, and uses the following packages: * MLBDD by Arlen Cox, * Menhir by François Pottier and Yann Régis-Gianas, * Pprint by François Pottier, * XML-Light by Nicolas Cannasse and Jacques Garrigue.

RELEASE FUNCTIONAL DESCRIPTION: Version 0.2: * MEL: ad hoc language for the declaration of variable dimension multi-mode DAE systems * automatic parsing, model checking and model allocation * XML output for the list of evaluation blocks (parameters, equations, unknowns to be computed) * new algorithms for the mode-dependent scheduling and the evaluation of the scheduling in a given mode

NEWS OF THE YEAR: It has been possible to perform the structural analysis of systems with more than 750 equations and 10 to the power 23 modes, therefore demonstrating the scalability of the method.

- Authors: Benoît Caillaud and Mathias Malandain
- Contact: Benoît Caillaud

6. New Results

6.1. Mathematical Foundations of Physical Systems Modeling Languages

Participants: Albert Benveniste, Benoît Caillaud, Mathias Malandain.

Modern modeling languages for general physical systems, such as Modelica or Simscape, rely on Differential Algebraic Equations (DAE), i.e., constraints of the form $f(\dot{x}, x, u) = 0$. This facilitates modeling from first principles of the physics. This year we completed the development of the mathematical theory needed to sound, on solid mathematical bases, the design of compilers and tools for DAE based physical modeling languages.

Unlike Ordinary Differential Equations (ODE, of the form $\dot{x} = g(x, u)$), DAE exhibit subtle issues because of the notion of *differentiation index* and related *latent equations*—ODE are DAE of index zero for which no latent equation needs to be considered. Prior to generating execution code and calling solvers, the compilation of such languages requires a nontrivial *structural analysis* step that reduces the differentiation index to a level acceptable by DAE solvers.

Multimode DAE systems, having multiple modes with mode-dependent dynamics and state-dependent mode switching, are much harder to deal with. The main difficulty is the handling of the events of mode change. Unfortunately, the large literature devoted to the numerical analysis of DAEs does not cover the multimode case, typically saying nothing about mode changes. This lack of foundations causes numerous difficulties to the existing modeling tools. Some models are well handled, others are not, with no clear boundary between the two classes. Basically, no tool exists that performs a correct structural analysis taking multiple modes and mode changes into account.

In our work, we developed a comprehensive mathematical approach supporting compilation and code generation for this class of languages. Its core is the *structural analysis of multimode DAE systems,* taking both multiple modes and mode changes into account. As a byproduct of this structural analysis, we propose well sound criteria for accepting or rejecting models at compile time.

For our mathematical development, we rely on *nonstandard analysis,* which allows us to cast hybrid systems dynamics to discrete time dynamics with infinitesimal step size, thus providing a uniform framework for handling both continuous dynamics and mode change events.

A big comprehensive document has been written, which will be finalized and submitted next year.

6.2. Structural analysis of multimode DAE systems

Participants: Albert Benveniste, Benoît Caillaud, Khalil Ghorbal, Mathias Malandain.

The Hycomes team has obtained two results related to the structural analysis of multimode DAE systems.

6.2.1. Impulsive behavior of multimode DAE systems

A major difficulty with multimode DAE systems are the commutations from one mode to another one when the number of equations may change and variables may exhibit impulsive behavior, meaning that not only the trajectory of the system may be discontinuous, but moreover, some variables may be Dirac measures at the instant of mode changes. In [7] , we compare two radically different approaches to the structural analysis problem of mode changes. The first one is a classical approach, for a restricted class of DAE systems, for which the existence and uniqueness of an impulsive state jump is proved. The second approach is based on nonstandard analysis and is proved to generalize the former approach, to a larger class of multimode DAE systems. The most interesting feature of the latter approach is that it defines the state-jump as the standardization of the solution of a system of system of difference equations, in the framework of nonstandard analysis.

6.2.2. An implicit structural analysis method for multimode DAE systems

Modeling languages and tools based on Differential Algebraic Equations (DAE) bring several specific issues that do not exist with modeling languages based on Ordinary Differential Equations. The main problem is the determination of the differentiation index and latent equations. Prior to generating simulation code and calling solvers, the compilation of a model requires a structural analysis step, which reduces the differentiation index to a level acceptable by numerical solvers.

The Modelica language, among others, allows hybrid models with multiple modes, mode-dependent dynamics and state-dependent mode switching. These Multimode DAE (mDAE) systems are much harder to deal with. The main difficulties are (i) the combinatorial explosion of the number of modes, and (ii) the correct handling of mode switchings.

The focus of the paper [31] is on the first issue, namely: How can one perform a structural analysis of an mDAE in all possible modes, without enumerating these modes? A structural analysis algorithm for mDAE systems is presented, based on an implicit representation of the varying structure of an mDAE. It generalizes J. Pryce's Σ-method [56] to the multimode case and uses Binary Decision Diagrams (BDD) to represent the mode-dependent structure of an mDAE. The algorithm determines, as a function of the mode, the set of latent equations, the leading variables and the state vector. This is then used to compute a mode-dependent block-triangular decomposition of the system, that can be used to generate simulation code with a mode-dependent scheduling of the blocks of equations.

This method has been implemented in the IsamDAE software. This has allowed the Hycomes team to evaluate the performance and scalability of the method on several examples. In particular, it has been possible to perform the structural analysis of systems with more than 750 equations and 10^{23} modes.

6.3. Functional Decision Diagrams: A Unifying Data Structure For Binary Decision Diagrams

Participants: Joan Thibault, Khalil Ghorbal.

Zero-suppressed binary Decision Diagram (ZDD) is a notable alternative data structure of Reduced Ordered Binary Decision Diagram (ROBDD) that achieves a better size compression rate for Boolean functions that evaluate to zero almost everywhere. Deciding *a priori* which variant is more suitable to represent a given Boolean function is as hard as constructing the diagrams themselves. Moreover, converting a ZDD to a ROBDD (or vice versa) often has a prohibitive cost. This observation could be in fact stated about almost all existing BDD variants as it essentially stems from the non-compatibility of the reduction rules used to build such diagrams. Indeed, they are neither interchangeable nor composable. In [8], we investigate a novel functional framework, termed Lambda Decision Diagram (LDD), that ambitions to classify the already existing variants as implementations of special LDD models while suggesting, in a principled way, new models that exploit application-dependant properties to further reduce the diagram's size. We show how the reduction rules we use locally capture the global impact of each variable on the output of the entire function. Such knowledge suggests a variable ordering that sharply contrasts with the static fixed global ordering in the already existing variants as well as the dynamic reordering techniques commonly used.

7. Bilateral Contracts and Grants with Industry

7.1. Glose: Globalisation for Systems Engineering

Participants: Benoît Caillaud, Benoît Vernay.

Glose is a bilateral collaboration between Inria and Safran Tech., the corporate research entity of Safran Group. It started late 2017 for a duration of 44 months. Three Inria teams are involved in this collaboration: Diverse (Inria Rennes), Hycomes and Kairos (Inria Sophia-Antipolis). The scope of the collaboration is systems engineering and co-simulation.

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. 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?

Co-simulation requires that models are provided with interfaces, specifying static and dynamic properties about the model and its expected environments. Interfaces are required to define how each model may synchronize and communicate, and how the model should be used. For instance, an interface should define (i) which variables are inputs, which are outputs, (ii) their data types, physical units, and sampling periods, but also (iii) the environmental assumptions under which the model is valid, and (iv) the causal dependencies between input and output variables and for continuous-time models, (v) the stiffness of the model, often expressed as a time-varying Jacobian matrix.

Formally, an interface is an abstraction of a model's behavior. A typical example of interface formalism for 0D continuous-time models is the FMI standard. Co-simulation also requires that a model of the system architecture is provided. This architectural model specifies how components are interconnected, how they communicate and how computations are scheduled. This is not limited to the topology of the architecture, and should also specify how components interact. For instance, variables in continuous-time models may have different data-types and physical units. Conversion may be required when continuous-time models are plugged together. Another fine example is the coupling of a 3D finite-element model to a 0D model: effort and flow fields computed in the 3D model must be averaged in a scalar value, before it can be sent to the 0D model, and conversely, scalar values computed by the 0D model must be distributed as a (vector) field along a boundary manifold of the 3D model. For discrete-time models (eg., software), components may communicate in many ways (shared variables, message passing, ...), and computations can be time- or event-triggered. All these features are captured as data-/behavior-coordination patterns, as exemplified by the GEMOC initiative [0].

In the Glose project, we propose to formalize the behavioral semantics of several 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 co-simulation of a system-level architecture consists in an orchestration of hundreds to thousands of components. This orchestration is achieved by a master algorithm, in charge of triggering the communication and computation steps of each component. It takes into account the components' interfaces, and the data-/behavior-coordination patterns found in the system architecture model. Because simulation scalability is a major issue, the scheduling policy computed by the master algorithm should be optimal. Parallel or distributed simulations may even be required. This implies that the master algorithm should be hierarchical and possibly distributed.

In 2019, the Hycomes team has been working on the use of Quantized State System (QSS) nethods for the cosimulation of aeronautics system models. The aim is to design new distributed simulation protocols, capable of simulating large, but heterogeneous system models. The investigation is on the trade-offs between pessimistic simulation techniques, where no roll-back is required, and speculative methods, where roll-back may be required. The latter method can be beneficial to the performance and scalability of the simulation, provided roll-backs do not happen too often. The models under consideration are cyberphysical systems consisting in both Modelica models (for the physics) and discrete-time models expressed in a dedicated language (for the control).

[0] http://gemoc.org

In 2019, the Hycones team has delivered one report, detailing the state-of-the-art techniques for continuous systems cosimulation.

8. Partnerships and Cooperations

8.1. Regional Initiatives

Participants: Benoît Caillaud, Aurélien Lamercerie.

The Hycomes has been participating to the SUNSET project (2016–2019) of the CominLabs excellence laboratory [0]. 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. The main contribution of the Hycomes team to this project has been the development of Demodocos, a process model synthesis tool, capable of generating models of a surgical procedure, from a few recordings of actual procedures. Demodocos has been interfaced to the `#SEVEN` virtual reality scenario modeling language and engine, developed in the Hybrid team at Inria Rennes. In 2019, the team has contributed to two publications presenting experimental results of the SUNSET project [9][6].

8.2. National Initiatives

8.2.1. 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 Bernard Brogliato Alexandre Rocca	Tripop	Inria Grenoble Rhône Alpes
Albert Benveniste Benoît Caillaud Khalil Ghorbal Christelle Kozaily Mathias Malandain Benoît Vernay	Hycomes	Inria Rennes Bretagne Atlantique
Marc Pouzet Tim Bourke Imsail Lakhim-Bennani	Parkas	ENS & Inria Paris
Goran Frehse	SSH	ENSTA Paris-Tech.
Antoine Girard		L2S-CNRS, Saclay
Eric Goubault Sylvie Putot	Cosynus	LIX, École Polytechnique, 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).

[0] http://www.s3pm.cominlabs.ueb.eu/

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.

In 2019, three general meetings have been organized, with presentations of the partners on new results related to hybrid systems modeling and verification.

Two PhDs are funded by the ModeliScale IPL. Both started in October 2018:

- Christelle Kozaily has started a PhD, under the supervision of Vincent Acary (TRIPOP team at Inria Grenoble), Benoît Caillaud, Khalil Ghorbal on the structural and numerical analysis of non-smooth DAE systems. She is located in the Hycomes team at Inria Rennes.
- Ismail Lahkim-Bennani has started a PhD under the supervision of Goran Frehse (ENSTA Paris-Tech.) and Marc Pouzet (PARKAS team, Inria/ENS Paris). His PhD topic is on random testing of hybrid systems, using techniques inspired by QuickCheck [36].

8.2.2. FUI ModeliScale: Scalable Modeling and Simulation of Large Cyber-Physical Systems

Participants: Albert Benveniste, Benoît Caillaud, Khalil Ghorbal, Mathias Malandain.

FUI ModeliScale is a French national collaborative project coordinated by Dassault Systèmes. The partners of this project are: EDF and Engie as main industrial users; DPS, Eurobios and PhiMeca are SME providing mathematical modeling expertise; CEA INES (Chambéry) and Inria are the academic partners. The project started January 2018, for a maximal duration of 42 months. Three Inria teams are contributing to the project : Hycomes, Parkas (Inria Paris / ENS) and Tripop (Inria Grenoble / LJK).

The focus of the project is on the scalable analysis, compilation and simulation of large Modelica models. One of the main contributions expected from Inria are:

- A novel structural analysis algorithms for multimode DAE systems, capable of handling large systems of guarded equations, that do not depend on the enumeration of a possibly exponential number of modes.
- The partitioning and high-performance distributed co-simulation of large Modelica models, based on the results of the structural analysis.

In 2019, the effort has been put on the first objective, and two important milestones have been reached:

- The design of a novel algorithm for the structural analysis of multimode DAE systems. This algorithm is a generalization of the Pryce structural analysis method to the multimode case. The key feature of our method is that it works on implicit representations of the set of modes, and of the varying structure of the multimode DAE. In other words, it does not imply the enumeration of the system's modes. Performing the structural analysis at compile-time brings two decisive advantages: 1/ it allows to deliver to the user precise diagnostics about the model, and can be compared type-checking in programming languages; 2/ it is instrumental for the generation of efficient simulation code. Our algorithm is the first method enabling the compile-time analysis of systems with extremely large combinatorics of modes.
- Our multimode DAE structural analysis algorithm has been implemented in IsamDAE, a software comprizing an algorithmic library, to be used in modeling language compilers (Modelica tools) and a standalone tool, to be used independently of a complex Modelica toolset. IsamDAE has allowed to benchmark the method against several families of models, inspired by case-studies developed by industrial partners of the FUI ModeliScale project. Despite the tool is still under development, we have already been able to deal with models with up to 10^{23} modes.

On top of these two main results, the Hycomes team has started investigating the use of Quantized Space Systems (QSS), for the simulation of large DAE systems. QSSs simulation (QSS) was introduced in the early 2000's by F. Cellier and E. Kofman as an alternative to time-based simulation, which is the dominant approach to ODE/DAE systems simulation. Rather than linking QSS to Discrete Event Simulation, we propose to relate it to Synchronous Programming and its continuous time extension Zelus. In the deliverable [20], we expose our understanding of QSS and its variants, then we propose ideas toward a QSS-based cosimulation, by building on top of our knowledge on distributed executions of synchronous programs.

The plan for 2020 is to extend our structural analysis to cover impulsive mode changes and the consistent initialization problem, in the multimode case. A coupling of IsamDAE with Dymola (Dassault Système's commercial implementation of the Modelica language) is under development.

Another future development is to turn our structural analysis method to a compositional method, where large models could be considered by parts. This is a key problem in the Modelica language, as the compilation of a Modelica model is not modular.

Work on QSS methods will continue, and we envision to prototype a QSS-based distributed simulation method for hybrid ODE systems, based on the Zélus language.

8.3. International Initiatives

8.3.1. Inria International Partners

8.3.1.1. Informal International Partners

We have a long standing informal collaboration with Martin Otter (DLR, Munich, Germany) and Hilding Helmqvist (Mogram AB, Lund, Sweden). In 2019, this fruitful collaboration has resulted in one publication [7]. The publication draws links between two radically different, but equivalent approaches to the same problem: the impulsive behavior of some multimode DAE, when it is switching from one mode to another. The first approach relies on a transformation of the multimode DAE system to a special index one form, for which state-jumps are proved to be solution of a system of algebraic equations relating right limits to left limits. The second approach builds on the use of nonstandard analysis, combined with the heritage of synchronous programming languages, particularly on the concept of constructive semantics. This gives a formulation of the state-jumps, as a system of difference equations, with an infinitesimal time-step. The latter approach is more general than the former, in the sense that impulsive behavior can be characterized for a larger class of multimode DAE systems. Yet, both approaches coincide on a restricted class of multimode DAEs.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Selection

9.1.1.1. Member of the Conference Program Committees

- Albert Benveniste has served on the Program Committee of the Modelica Conference 2019.
- Khalil Ghorbal has served on the Program Committee of the VMCAI 2019 conference.
- Benoît Caillaud has served on the Steering Committee of the ACSD 2019 conference.

9.1.2. Invited Talks

Albert Benveniste has given an invited talk titled “Why considering nonstandard semantics for hybrid systems and how to reconcile it with superdense time semantics” at the Oded Maler Memorial workshop at the HSCC'19 conference in Montreal, Canada.

9.1.3. Leadership within the Scientific Community

Albert Benveniste is member of the French Académie des Technologies.

9.1.4. Scientific Expertise

- In 2019, Benoît Caillaud has reviewed collaborative research project proposals submitted to the French national funding agency ANR. As an Inria Evaluation Committee member, he has served on several Inria hiring and promotion committees.
- Albert Benveniste is president of the Scientific Council of Orange and member of the Scientific Council of Safran.

9.1.5. Research Administration

- Albert Benveniste is member of the Burex (Executive Bureau) of the Cominlabs Labex [0].
- Benoît Caillaud is in charge of the IPL ModeliScale [0] national initiative funded by Inria. He is also head of the Programming Languages & Software Engineering department [0] of IRISA.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

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

9.2.2. Supervision

PhD: Christelle Kozaily, Structural analysis of nonsmooth dynamical systems, university of Rennes 1, co-supervised by Vincent Acary (Tripop [0] team at Inria Grenoble), Benoît Caillaud and Kahlil Ghorbal, started October 2018.

PhD: Aurélien Lamercerie, Formal analysis of cyber-physical systems requirements expressed in natural language, university of Rennes 1, co-supervised by par Benoît Caillaud et Annie Forêt (SemLIS [0] team of IRISA), started December 2017.

PhD: Joan Thibault, Structural Analysis Techniques for Binary Decision Diagrams, university of Rennes 1, co-supervised by Benoît Caillaud and Khalil Ghorbal.

9.2.3. Juries

Benoît Caillaud has served as president of the jury for Hugo Bazille's PhD defense, at the University of Rennes 1.

9.3. Popularization

9.3.1. Internal action

The Hycomes team has hosted short internships for secondary school students. This has been an opportunity to promote women in computing, since three female students visited the team for four days, to discover what scientific research is, and what research in computer science could mean. All team members contributed to the initiative.

[0] https://cominlabs.u-bretagneloire.fr/governance
[0] https://team.inria.fr/modeliscale/
[0] http://www.irisa.fr/en/departments/d4-language-and-software-engineering
[0] https://team.inria.fr/tripop/
[0] https://www-semlis.irisa.fr

10. Bibliography

Major publications by the team in recent years

[1] A. Benveniste, T. Bourke, B. Caillaud, J.-L. Colaço, C. Pasteur, M. Pouzet.*Building a Hybrid Systems Modeler on Synchronous Languages Principles*, in "Proceedings of the IEEE", September 2018, vol. 106, n^{o} 9, p. 1568–1592 [*DOI :* 10.1109/JPROC.2018.2858016], https://hal.inria.fr/hal-01879026

[2] 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^{o} 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

[3] A. Benveniste, B. Caillaud, D. Nickovic, R. Passerone, J.-B. Raclet, P. Reinkemeier, A. Sangiovanni-Vincentelli, W. Damm, T. Henzinger, K. G. Larsen.*Contracts for System Design*, in "Foundations and Trends in Electronic Design Automation", 2018, vol. 12, n^{o} 2-3, p. 124-400 [*DOI :* 10.1561/1000000053], https://hal.inria.fr/hal-01971429

[4] J.-B. Jeannin, K. Ghorbal, Y. Kouskoulas, A. Schmidt, R. Gardner, S. Mitsch, A. Platzer.*A Formally Verified Hybrid System for Safe Advisories in the Next-Generation Airborne Collision Avoidance System*, in "International Journal on Software Tools for Technology Transfer", November 2017, vol. 19, n^{o} 6, p. 717-741 [*DOI :* 10.1007/s10009-016-0434-1], https://hal.archives-ouvertes.fr/hal-01232365

[5] 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^{o} 5s, p. 185:1–185:19 [*DOI :* 10.1145/3126506], https://hal.inria.fr/hal-01658196

Publications of the year

Articles in International Peer-Reviewed Journal

[6] M.-S. Bracq, E. Michinov, B. Arnaldi, B. Caillaud, B. Gibaud, V. Gouranton, P. Jannin.*Learning procedural skills with a virtual reality simulator An acceptability study*, in "Nurse Education Today", August 2019, vol. 79, p. 153-160 [*DOI :* 10.1016/j.nedt.2019.05.026], https://hal-univ-rennes1.archives-ouvertes.fr/hal-02150192

Scientific Books (or Scientific Book chapters)

[7] A. Benveniste, B. Caillaud, H. Elmqvist, K. Ghorbal, M. Otter, M. Pouzet.*Multi-Mode DAE Models - Challenges, Theory and Implementation*, in "Computing and Software Science: State of the Art and Perspectives", Lecture Notes in Computer Science, Springer, October 2019, vol. 10000, p. 283-310 [*DOI :* 10.1007/978-3-319-91908-9_16], https://hal.inria.fr/hal-02333603

Research Reports

[8] J. Thibault, K. Ghorbal.*Functional Decision Diagrams: A Unifying Data Structure For Binary Decision Diagrams*, Inria Rennes - Bretagne Atlantique and University of Rennes 1, France, November 2019, n^{o} RR-9306, https://hal.inria.fr/hal-02369112

Other Publications

[9] M.-S. BRACQ, E. MICHINOV, B. ARNALDI, A. AUDINOT, B. CAILLAUD, B. GIBAUD, V. GOURANTON, P.-L. HENAUX, A. LAMERCERIE, B. NOGUES, P. JANNIN.*Modeling, Simulation and Training Procedural Skills: User experience and acceptability of a virtual reality simulator for scrub nurses in neurosurgery*, January 2019, vol. 36, 1, 19th International Meeting on Simulation in Healthcare (IMSH), Poster, https://hal.archives-ouvertes.fr/hal-02123682

[10] A. LAMERCERIE.*Une algèbre des automates d'acceptation propositionnelle déterministes comme théorie d'interface pour la conception de systèmes cyberphysiques*, November 2019, 1, MSR 2019 - 12ème Colloque sur la Modélisation des Systèmes Réactifs, Nov 2019, Angers, France, Poster, https://hal.archives-ouvertes.fr/hal-02432696

References in notes

[11] N. J. CUTLAND (editor). *Nonstandard analysis and its applications*, Cambridge Univ. Press, 1988

[12] *IEEE Standard VHDL Analog and Mixed-Signal Extensions, Std 1076.1-1999*, 1999, http://dx.doi.org/10.1109/IEEESTD.1999.90578

[13] 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^o 94

[14] C. BAIER, J.-P. KATOEN.*Principles of Model Checking*, MIT Press, Cambridge, 2008

[15] A. BENVENISTE, T. BOURKE, B. CAILLAUD, J.-L. COLAÇO, C. PASTEUR, M. POUZET.*Building a Hybrid Systems Modeler on Synchronous Languages Principles*, in "Proceedings of the IEEE", September 2018, vol. 106, n^o 9, p. 1568–1592 [*DOI :* 10.1109/JPROC.2018.2858016], https://hal.inria.fr/hal-01879026

[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, 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

[19] 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

[20] A. BENVENISTE, M. POUZET, M. MALANDAIN.*Bibliography report on Quantized Space Systems simulation — Proposals for QSS-based co-simulation of large DAE systems*, Inria, January 2020, n^{o} M2.2.2_1

[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^{o} 2, p. 251–274

[27] G. BOUDOL, K. G. LARSEN.*Graphical Versus Logical Specifications*, in "Theor. Comput. Sci.", 1992, vol. 106, n^{o} 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^{o} 34, p. 4373-4404 [*DOI :* 10.1016/J.TCS.2011.05.010], http://hal.inria.fr/hal-00654003/en

[31] B. CAILLAUD, M. MALANDAIN, J. THIBAULT.*Implicit Structural Analysis of Multimode DAE Systems*, in "23rd ACM International Conference on Hybrid Systems: Computation and Control (HSCC 2020)", Sydney, Australia, April 2020, to appear

[32] S. L. CAMPBELL, C. W. GEAR.*The index of general nonlinear DAEs*, in "Numerische Mathematik", dec 1995, vol. 72, n^{o} 2, p. 173–196, http://dx.doi.org/10.1007/s002110050165

[33] 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

[34] 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

[35] 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

[36] K. CLAESSEN, J. HUGHES.*QuickCheck: a lightweight tool for random testing of Haskell programs*, in "Proceedings of the Fifth ACM SIGPLAN International Conference on Functional Programming (ICFP '00), Montreal, Canada, September 18-21, 2000.", M. ODERSKY, P. WADLER (editors), ACM, 2000, p. 268–279, https://doi.org/10.1145/351240.351266

[37] E. CLARKE, O. GRUMBERG, D. PELED.*Model Checking*, MIT Press, 1999

[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. WASOWSKI.*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] J. EDMONDS, R. M. KARP.*Theoretical improvements in algorithmic efficiency for network flow problems*, in "Journal of the ACM", 1972, vol. 19, n^{o} 2, p. 248–264, http://dx.doi.org/10.1145/321694.321699

[44] A. V. GOLDBERG, R. E. TARJAN.*A new approach to the maximum flow problem*, in "Proceedings of the eighteenth annual ACM symposium on Theory of computing (STOC'86)", 1986, http://dx.doi.org/10.1145/12130.12144

[45] 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

[46] L. LAMPORT.*Proving the Correctness of Multiprocess Programs*, in "IEEE Trans. Software Eng.", 1977, vol. 3, n^{o} 2, p. 125-143

[47] 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

[48] 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

[49] T. LINDSTRØM.*An Invitation to Nonstandard Analysis*, in "Nonstandard Analysis and its Applications", N. J. CUTLAND (editor), Cambridge Univ. Press, 1988, p. 1–105

[50] N. A. LYNCH.*Input/Output Automata: Basic, Timed, Hybrid, Probabilistic and Dynamic*, in "CONCUR", R. M. AMADIO, D. LUGIEZ (editors), Lecture Notes in Computer Science, Springer, 2003, vol. 2761, p. 187-188

[51] N. A. LYNCH, E. W. STARK.*A Proof of the Kahn Principle for Input/Output Automata*, in "Inf. Comput.", 1989, vol. 82, n^{o} 1, p. 81-92

[52] Z. MANNA, A. PNUELI.*Temporal verification of reactive systems: Safety*, Springer, 1995

[53] B. MEYER.*Applying "Design by Contract"*, in "Computer", October 1992, vol. 25, n^{o} 10, p. 40–51, http://dx.doi.org/10.1109/2.161279

[54] 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^{o} 12, p. 3329–3345

[55] C. PANTELIDES.*The consistent initialization of differential-algebraic systems*, in "SIAM J. Sci. Stat. Comput.", 1988, vol. 9, n^{o} 2, p. 213–231

[56] J. D. PRYCE.*A Simple Structural Analysis Method for DAEs*, in "BIT Numerical Mathematics", March 2001, vol. 41, n^{o} 2, p. 364–394, http://dx.doi.org/10.1023/a:1021998624799

[57] 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^{o} 1-2, p. 119-149 [*DOI :* 10.3233/FI-2011-416], http://hal.inria.fr/inria-00554283/en

[58] A. ROBINSON.*Non-Standard Analysis*, Princeton Landmarks in Mathematics, 1996, ISBN 0-691-04490-2

[59] 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

[60] 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

[61] 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

[62] 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

[63] 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

Université Gustave Eiffel

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Optimization and control of dynamic systems

Table of contents

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. Team, Visitors, External Collaborators

Research Scientists

Laurent Mevel [Team leader, Inria, Senior Researcher, HDR]
Michael Doehler [Inria, Researcher]
Jean Dumoulin [IFSTTAR, Researcher]
Xavier Chapeleau [IFSTTAR, Researcher]
Qinghua Zhang [Inria, Senior Researcher, HDR]

Faculty Member

Frederic Gillot [Ecole centrale de Lyon, Associate Professor, until Aug 2019]

Technical Staff

Szymon Gres [Inria, Engineer, Carnot, from Oct 2019]
Ivan Guéguen [IFSTTAR, Engineer]
Vincent Le Cam [IFSTTAR, Engineer]
Mathieu Le Pen [IFSTTAR, Engineer, until Sep 2019]
Arthur Bouché [IFSTTAR, Engineer, since 2019]

PhD Students

Antoine Bassil [IFSTTAR, until November 2019]
David Pallier [IFSTTAR, IETR]
Thibaud Toullier [Inria, until Oct 2019]
Bian Xiong [Inria]
Quentin Sourisseau [Veritas]
Eva Viefhues [BAM]
Alexander Mendler [University of British Columbia]

Post-Doctoral Fellows

Ludovic Gaverina [IFSTTAR]
Nicolas Le Touz [IFSTTAR]

Visiting Scientist

Wentao Tang [Harbin Institut of Technology (China),PhD student, from May 2019]

2. Overall Objectives

2.1. Overall Objectives

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,

$$\left\{ \begin{array}{rcl} X_{n+1} & = & AX_n + V_n \\ Y_n & = & HX_n + W_n \end{array} \right. \tag{3}$$

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)

$$M\ddot{Z} + C\dot{Z} + KZ = V \tag{4}$$

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).

$$M\ddot{Z} + C\dot{Z} + KZ = U^2 DZ + UE\dot{Z} + V \tag{5}$$

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 $k = 0, 1, \cdots, P$ to get the data matrix $\mathcal{Y}_+$ and stack the column vectors Y_k^T for $k = 0, -1, \cdots, -P$ 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 ε; tuning ε 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 θ_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 subspace-based methods, for both the identification and the monitoring of the eigenstructure (λ, ϕ_λ) of the state transition matrix F of a linear dynamical state-space system :

$$\left\{\begin{array}{rcl} X_{k+1} & = & F\ X_k + V_{k+1} \\ Y_k & = & H\ X_k + W_k \end{array}\right. , \tag{6}$$

namely the $(\lambda, \varphi_\lambda)$ defined by :

$$\det\ (F - \lambda\ I) = 0, \quad (F - \lambda\ I)\ \phi_\lambda = 0, \quad \varphi_\lambda \triangleq H\ \phi_\lambda \tag{7}$$

The (canonical) parameter vector in that case is :

$$\theta \triangleq \begin{pmatrix} \Lambda \\ \mathrm{vec}\Phi \end{pmatrix} \tag{8}$$

where Λ is the vector whose elements are the eigenvalues λ, Φ is the matrix whose columns are the φ_λ'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 [51].

Let $R_i \triangleq \mathbf{E}\left(Y_k\ Y_{k-i}^T\right)$ and:

$$\mathcal{H}_{p+1,q} \triangleq \begin{pmatrix} R_1 & R_2 & \vdots & R_q \\ R_2 & R_3 & \vdots & R_{q+1} \\ \vdots & \vdots & \vdots & \vdots \\ R_{p+1} & R_{p+2} & \vdots & R_{p+q} \end{pmatrix} \triangleq \mathrm{Hank}\,(R_i) \tag{9}$$

be the output covariance and Hankel matrices, respectively; and: $G \triangleq \mathbf{E}\left(X_k Y_{k-1}^T\right)$. Direct computations of the R_i's from the equations (4) lead to the well known key factorizations :

$$\begin{aligned} R_i &= HF^{i-1}G \\ \mathcal{H}_{p+1,q} &= \mathcal{O}_{p+1}(H,F)\ \mathcal{C}_q(F,G) \end{aligned} \tag{10}$$

where:

$$\mathcal{O}_{p+1}(H,F) \triangleq \begin{pmatrix} H \\ HF \\ \vdots \\ HF^p \end{pmatrix} \quad \text{and} \quad \mathcal{C}_q(F,G) \triangleq (G\ FG\ \cdots\ F^{q-1}G) \tag{11}$$

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 (λ, ϕ_λ) 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 θ_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:

$$\mathcal{O}_{p+1}(\theta) = \begin{pmatrix} \Phi \\ \Phi\Delta \\ \vdots \\ \Phi\Delta^p \end{pmatrix} \tag{12}$$

where $\Delta \stackrel{\Delta}{=} \mathrm{diag}(\Lambda)$, and Λ and Φ are as in (6). Whether a nominal parameter θ_0 fits a given output covariance sequence $(R_j)_j$ is characterized by:

$$\mathcal{O}_{p+1}(\theta_0) \text{ and } \mathcal{H}_{p+1,q} \text{ have the same left kernel space.} \tag{13}$$

This property can be checked as follows. From the nominal θ_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:

$$U^T\, U = I_s \text{ and } U^T\, \mathcal{O}_{p+1}(\theta_0) = 0 \tag{14}$$

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 θ_0. Then the characterization writes:

$$U(\theta_0)^T\, \mathcal{H}_{p+1,q} = 0 \tag{15}$$

3.1.2.1. Residual associated with subspace identification.

Assume now that *a reference* θ_0 *and a new sample* $Y_1, \cdots, Y_N$ *are available.* For checking whether the data agree with θ_0, the idea is to compute the empirical Hankel matrix $\widehat{\mathcal{H}}_{p+1,q}$:

$$\widehat{\mathcal{H}}_{p+1,q} \stackrel{\Delta}{=} \mathrm{Hank}\left(\widehat{R}_i\right), \quad \widehat{R}_i \stackrel{\Delta}{=} 1/(N-i) \sum_{k=i+1}^{N} Y_k\, Y_{k-i}^T \tag{16}$$

and to define the residual vector:

$$\zeta_N(\theta_0) \stackrel{\Delta}{=} \sqrt{N}\ \mathrm{vec}\left(U(\theta_0)^T\, \widehat{\mathcal{H}}_{p+1,q}\right) \tag{17}$$

Let θ 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 θ. From (13), we know that $\zeta_N(\theta_0)$ has zero mean when no change occurs in θ, 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 [45] that the residual is asymptotically Gaussian :

$$\zeta_N \xrightarrow[N \to \infty]{} \begin{cases} \mathcal{N}(0, \Sigma) & \text{under } \mathbf{P}_{\theta_0} , \\ \\ \mathcal{N}(\mathcal{J}\,\eta, \Sigma) & \text{under } \mathbf{P}_{\theta_0+\eta/\sqrt{N}} , \end{cases} \tag{18}$$

where the asymptotic covariance matrix Σ 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 χ^2-test, provided that $\mathcal{J}$ is full rank and Σ is invertible :

$$\chi^2 = \overline{\zeta}^{\,T}\, \mathbf{F}^{-1}\, \overline{\zeta} \gtrless \lambda \ , \tag{19}$$

where

$$\overline{\zeta} \triangleq \mathcal{J}^T\, \Sigma^{-1}\, \zeta_N \quad \text{and} \quad \mathbf{F} \triangleq \mathcal{J}^T\, \Sigma^{-1}\, \mathcal{J} \ . \tag{20}$$

3.1.3. Diagnostics

A further monitoring step, often called *fault isolation*, consists in determining which (subsets of) components of the parameter vector θ 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 θ have changed ?*, can be addressed using either nuisance parameters elimination methods or a multiple hypotheses testing approach [44].

In most SHM applications, a complex physical system, characterized by a generally non identifiable parameter vector Φ has to be monitored using a simple (black-box) model characterized by an identifiable parameter vector θ. 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 θ and diagnosis in terms of the parameter vector Φ 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 Φ_0 of the application $\Phi \mapsto \theta(\Phi)$, and to use the sensitivity test for the components of the parameter vector Φ. Typically this results in the following type of directional test :

$$\chi^2_\Phi = \zeta^T\, \Sigma^{-1}\, \mathcal{J}\, \mathcal{J}_{\Phi\theta}\, (\mathcal{J}^T_{\Phi\theta}\, \mathcal{J}^T\, \Sigma^{-1}\, \mathcal{J}\, \mathcal{J}_{\Phi\theta})^{-1}\, \mathcal{J}^T_{\Phi\theta}\, \mathcal{J}^T\, \Sigma^{-1}\, \zeta \gtrless \lambda \ . \tag{21}$$

It should be clear that the selection of a particular parameterization Φ 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 introduces 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.

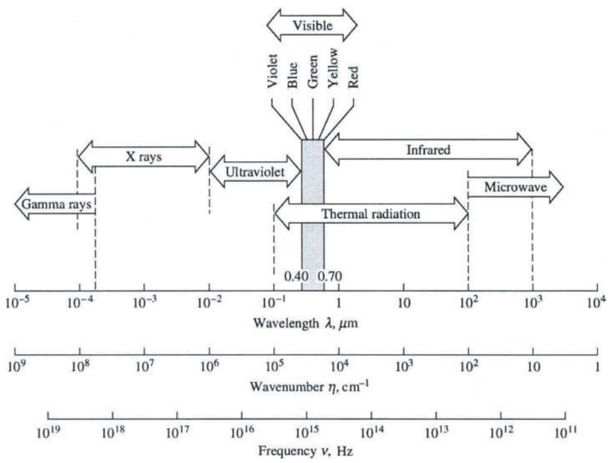


Figure 1. Electromagnetic spectrum - Credit MODEST, M.F. (1993). Radiative Heat Transfer. Academic Press.

For scientific purposes, infrared can be divided in three ranges of wavelength in which the application varies, see Table 1.

Table 1. Wavelength bands in the infrared according to ISO 20473:2007

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

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).

$$M^o_{\lambda,T} = \frac{C_1 \lambda^{-5}}{\exp^{\frac{C_2}{\lambda T}} - 1} \tag{22}$$

With λ the wavelength in m and T as the temperature in Kelvin. The C_1 and C_2 constants, respectively in W.m^2 and m.K are defined as follow:

$$\begin{aligned} C_1 &= 2hc^2\pi \\ C_2 &= h\frac{c}{k} \end{aligned} \tag{23}$$

with

- c, the electromagnetic wave speed (in vacuum c is the light speed in m.s^{-1}).
- $k = 1.381e^{-23}$ J.K^{-1} 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.
- $h \approx 6,62606957e^{-34}$ J.s The Plank constant. It is the link between the photons energy and their frequency.

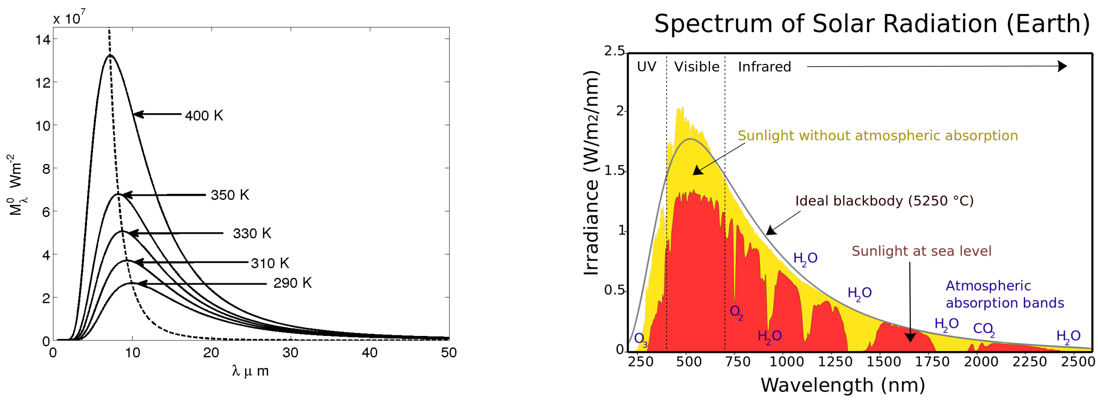


Figure 2. Left: Plank's law at various temperatures - Right: Energy spectrum of the atmosphere

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 depending on 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 purposes, 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 Ω. 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).

$$\varphi = k\nabla T \quad X \in \Omega \tag{24}$$

Where k is the thermal conductivity in W.m^{-1}.K ^{o}, ∇ is the gradient operator and φ is the heat flux density in Wm^{-2}. 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 yields to the expression of the heat conduction in all point of the domain Ω, cf Equation (23). This equation has been proposed by Joseph Fourier in 1811.

$$\rho C \frac{\partial T(X,t)}{\partial t} = \nabla \cdot (k\nabla T) + P \quad X \in \Omega \tag{25}$$

With $\nabla.()$ the divergence operator, C the specific heat capacity in J.kg^{-1}.^{o}K^{-1}, ρ the volumetric mass density in kg. m^{-3}, X the space variable $X = \{x, y, z\}$ and P a possible internal heat production in W.m^{-3}.
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).

$$\varphi = k\nabla T \cdot n = \underbrace{h\left(T_{fluid} - T_{Boundarie}\right)}_{\text{Convection}} + \underbrace{\epsilon\sigma_s\left(T^4_{environement} - T^4_{Boundary}\right)}_{\text{Radiation}} + \varphi_0 \quad X \in \partial\Omega \tag{26}$$

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^{-2}.K^{-1} and φ_0 an external energy contribution W.m^{-2}, in cases where the external energy contribution is artificial and controlled we call it active thermography (spotlight etc...), otherwise 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, model inversion is used, 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 yields another vector b, cf Equation (25). For example if A represents the heat transfer, b can be the temperature evolution.

$$AP = b \tag{27}$$

With A a matrix of size $n \times m$, P a vector of size m and b of size n, preferentially $n >> P$. 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).

$$P = A^{-1}b \tag{28}$$

Here we want to find the solution AP which is closest to the acquired measures M, Equation (27).

$$AP \approx \mathcal{M} \tag{29}$$

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.

$$min_P \left(\|AP - \mathcal{M}\|^2 \right) = min_P \left(\mathcal{F} \right) \tag{30}$$

Where $\mathcal{F}$ can be a product of matrix.

$$\mathcal{F} = [AP - \mathcal{M}]^T \, [AP - \mathcal{M}]$$

In some cases 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.

$$\nabla \mathcal{F}(P) = 2 \left[-\frac{\partial AP^T}{\partial P} \right] [AP - \mathcal{M}] = 2J(P)^T \, [AP - \mathcal{M}] \tag{31}$$

Where J is the sensitivity matrix of the model A with respect to the 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 the external heat flux has to be estimated, φ_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).

$$P^{k+1} = P^k + \left[(J^k)^T J^k + \mu^k \Omega^k \right]^{-1} (J^k)^T \left[\mathcal{M} - A(P^k) \right] \tag{32}$$

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], [50]. 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 [52], [48]. 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.

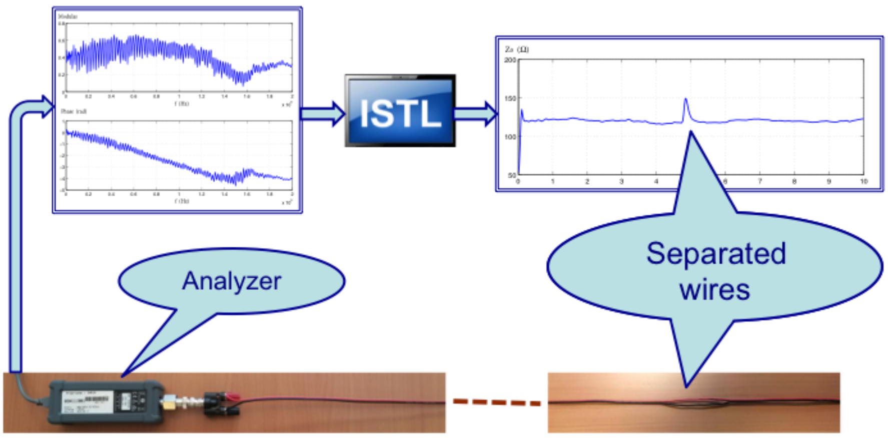


Figure 3. Inverse scattering software (ISTL) for cable soft fault diagnosis.

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 [49]

$$\begin{aligned} \frac{\partial}{\partial z}V(t,z) + L(z)\frac{\partial}{\partial t}I(t,z) + R(z)I(t,z) = 0 \\ \frac{\partial}{\partial z}I(t,z) + C(z)\frac{\partial}{\partial t}V(t,z) + G(z)V(t,z) = 0 \end{aligned} \tag{33}$$

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

$$V(t,a) = V_s(t) - R_sI(t,a). \tag{34}$$

At the right end of the cable (corresponding to $z = b$), the cable is connected to a load of impedance R_L, such that

$$V(t,b) \quad = R_LI(t,b). \tag{35}$$

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 ΔR_j, an inductance ΔL_j, a capacitance ΔC_j and a conductance ΔG_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 [47]. The visionary idea of applying this theory to solving the cable inverse problem goes also back to the 1980s [46]. After having completed some theoretic results directly linked to practice [14], [50], 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

$$\begin{aligned} \nu_1(\omega,x) = \frac{1}{2}\left[Z_0^{-\frac{1}{2}}(x)U(\omega,x) - Z_0^{\frac{1}{2}}(x)I(\omega,x)\right] \\ \nu_2(\omega,x) = \frac{1}{2}\left[Z_0^{-\frac{1}{2}}(x)U(\omega,x) + Z_0^{\frac{1}{2}}(x)I(\omega,x)\right] \end{aligned} \tag{36}$$

with

$$Z_0(x) = \sqrt{\frac{L(x)}{C(x)}}. \tag{37}$$

These transformations lead to the Zakharov-Shabat equations

$$\begin{aligned} \frac{d\nu_1(\omega,x)}{dx} + ik\nu_1(\omega,x) &= q^*(x)\nu_1(\omega,x) + q^+(x)\nu_2(\omega,x) \\ \frac{d\nu_2(\omega,x)}{dx} - ik\nu_2(\omega,x) &= q^-(x)\nu_1(\omega,x) - q^*(x)\nu_2(\omega,x) \end{aligned} \tag{38}$$

with

$$\begin{aligned} q^\pm(x) &= -\frac{1}{4}\frac{d}{dx}\left[ln\frac{L(x)}{C(x)}\right] \mp \frac{1}{2}\left[\frac{R(x)}{L(x)} - \frac{G(x)}{C(x)}\right] \\ &= -\frac{1}{2Z_0(x)}\frac{d}{dx}Z_0(x) \mp \frac{1}{2}\left[\frac{R(x)}{L(x)} - \frac{G(x)}{C(x)}\right] \\ q^*(x) &= \frac{1}{2}\left[\frac{R(x)}{L(x)} + \frac{G(x)}{C(x)}\right]. \end{aligned} \tag{39}$$

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 fiber-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 fiber-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.

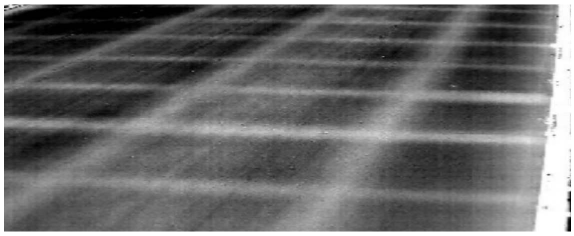


Figure 4. Left: Image in the visible spectrum of the deck surface - Right: PCT result on a bridge deck

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

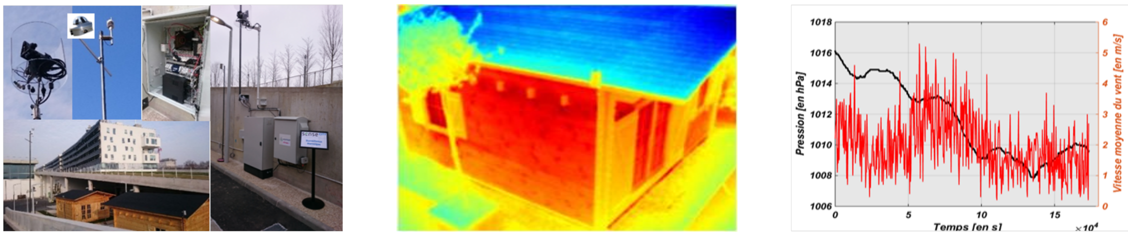


Figure 5. Various view and results of the SenseCity experimentation site - (site and hardware view, IR imaging, Environmental Monitoring)

project for IFSTTAR, which is a stakeholder in the Forever Open Road. Through its partnership with the COSYS (IFSTTAR) department, I4S is fully involved in the development of the 5th Generation Road.

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. Highlights of the Year

4.1. Highlights of the Year

4.1.1. Awards

- Our former PhD student Nicolas Le Touz received the Abertis Prize France for his thesis “Design and study of positive energy transport infrastructure: from thermomechanical modelling to the optimisation of such energy systems“, defended in November 2018. The Abertis Prize is awarded for research in transport infrastructure management.
- Nassif Berrabah, industrial PhD student of the I4S Team in collaboration with EDF, has defended his thesis on “Inverse problems for diagnosis of electric cables from reflectometry measurements” in November 2017. The research work of his thesis received the award of Scientific Prize from EDF R&D.

5. New Software and Platforms

5.1. Platforms

5.1.1. Pegase

PEGASE is the wireless platform developed by the team. The milestones for the PEGASE platfom in 2019 are

- finalization of the PEGASE3 hardware and software platform: SDK, decoding FPGA-based GPS frames,

- capacity (demonstrated during work with CEA Lost) of synchronous transmission / reception and ultrasonic wave phase.
- Writing documentation associated with the platform (User Guide ...).
- Start of the valuation process.

A prototype for modal analysis with several PEGASE platforms was developed. To obtain detailed modal information of large and very large structures, many sensors would be required to cover the geometry of the structure with a reasonable accuracy. However, when only a limited amount of sensors is available, large structures can be measured in several sensor setups, where some sensors remain fixed and some are moved between different measurement setups. With the sensors connected to different wireless platforms, the synchronous acquisition of data is required. A solution of data acquisition synchronization, as well as signal processing for merging the information taking into account the change of sensor positions and environmental variability has been developed and presented at IWSHM [31].

6. New Results

6.1. System identification

6.1.1. On Local LTI Model Coherence for LPV Interpolation

Participant: Qinghua Zhang.

In the local approach to linear parameter varying (LPV) system identification, it is widely acknowledged that locally estimated linear state-space models should be made coherent before being interpolated, but the accurate meaning of the term " coherent " or " coherence " is rarely defined. The purpose of this study is to analyze the relevance of two existing definitions and to point out the consequence of this analysis on the practice of LPV system identification. This work has been carried out in collaboration with Lennart Ljung of Linköping University and Rik Pintelon of Vriije Universiteit Brussel, and the results have been published in [22].

6.1.2. Stability Analysis of the Kalman Predictor

Participant: Qinghua Zhang.

The stability of the Kalman filter, though less often mentioned than the optimality in the recent literature, is a crucial property for real time applications. The purpose of this paper is to complete the classical stability analysis of the Kalman filter for general time varying systems. A proof of the stability of the one step ahead predictor, which is embedded in the Kalman filter, is presented in this paper, whereas the classical results were focused on the stability of the filter. The predictor stability is particularly important for linear parameter varying (LPV) system identification by means of prediction error minimization. This work has been carried out in collaboration with Liangquan Zhang of Beijing University of Posts and Telecommunications, and the results have been published in [23].

6.1.3. Regularized Adaptive Observer to Address Deficient Excitation

Participant: Qinghua Zhang.

Adaptive observers are recursive algorithms for joint estimation of both state variables and unknown parameters. Usually some persistent excitation (PE) condition is required for the convergence of adaptive observers. However, in practice, it may happen that the PE condition is not satisfied, because the available sensor signals do not contain sufficient information for the considered recursive estimation problem, which is ill-posed. To remedy the lack of PE condition, inspired by typical methods for solving ill-posed inverse problems, this paper proposes a regularized adaptive observer for general linear time varying (LTV) systems. Regularization terms are introduced in both state and parameter estimation recursions, in order to preserve the state-parameter decoupling transformation involved in the design of the adaptive observer. Like in typical ill-posed inverse problems, regularization implies an estimation bias, which can be reduced by using prior knowledge about the unknown parameters. This work has been carried out in collaboration with Fouad Giri and Tarek Ahmed-Ali of Université de Normandie, and the results have been presented at [40].

6.2. Damage detection and localization

6.2.1. Model sensitivity clustering for damage localization

Participants: Michael Doehler, Laurent Mevel.

The purpose of this paper is the development of a working damage localization method that is applicable on real data from complex structures. To achieve this goal, robust hypothesis tests are used, the sensitivity computation of the previously published residual is revisited for more precision thanks to reduced modal truncation errors, and an adequate clustering approach is proposed for the case of a high-dimensional FE parameterization for complex structures. Finally, an application of this framework is shown for the first time on experimental data for damage localization, namely in an ambient vibration test of a 3D steel frame at the University of British Columbia. [16]

6.2.2. Robustness to temperature changes for damage localization

Participants: Laurent Mevel, Michael Doehler, Alexander Mendler.

For structures in operation, temperature has been shown to be a major nuisance to the efficiency of such methods since the modal parameters are varying not only with damage but also due to temperature variations. For detection, environmental variation is hardly taken into account in localization approaches. In this paper, we propose a sensitivity-based correction of the identified modal parameters in the damaged state with respect to the temperature field in the reference state, based on a sensitivity analysis with respect to temperature dependent parameters of the finite element model in the reference state. The approach is then applied to the Stochastic Dynamic Damage Locating Vector (SDDLV) method, where its improved performance under non-uniform temperature variations is shown in a numerical application on a beam. [18], [26]

6.2.3. Robustness to temperature changes for damage detection

Participants: Laurent Mevel, Michael Doehler, Qinghua Zhang, Eva Viefhues.

Temperature affected vibration data is evaluated with a stochastic damage detection method, which relies on a null space based residual. A new approach is proposed, using model interpolation, where a global reference model is obtained from data in the reference state at several reference temperatures. Then, for a particular testing temperature, a local reference model is derived from the global reference model. Thus, a well fitting reference null space for the formulation of the residual is available when new data is tested for damage at arbitrary temperatures. Particular attention is paid to the computation of the residual's covariance, taking into account the uncertainty related to the null space estimate. This improves the test performance, resulting in a high probability of detection (PoD) of the new interpolation approach for global and local damages compared to previous approaches. [37]

6.3. Infrared Thermography

6.3.1. Long term thermal monitoring by standard passive Infrared thermography

Participants: Thibaud Toullier, Jean Dumoulin, Laurent Mevel.

The framework of latest technological improvements in low-cost infrared cameras have brought new opportunities for long-term infrastructures monitoring. Anyway, the accurate measurement of surfaces temperatures is facing the lack of knowledge of radiatives properties of the scene. By using multi-sensors instrumentation, the measurement model can be refined to get a better estimate of the temperature. To overcome a lack of sensors instrumentation, it has been shown that online and free available climatic data can be used. [15].

6.3.2. Long term thermal monitoring by multi-spectral infrared thermography

Participants: Thibaud Toullier, Jean Dumoulin, Laurent Mevel.

Bayesian methods to estimate simultaneously the emissivity and temperature have been developed and compared to literature's methods. A radiative exchange simulator of 3D scenes have been developed to compare those different methods on numerical data. This new software uses the hardware acceleration as well as a GPGPU approach to reduce the computation time. As a consequence, obtained numerical results emphasized an advanced use of multi-spectral infrared thermography for the monitoring of structures. This simultaneous estimation enables to have an estimate of the temperature by infrared thermography with a known uncertainty. [15].

6.4. Sensor and hardware based research

6.4.1. Fiber optic and interferometry

Participants: Xavier Chapeleau, Antoine Bassil.

The assessment of Coda Wave Interferometry (CWI) and Distributed Fiber Optics Sensing (DFOS) techniques for the detection of damages in a laboratory size reinforced concrete beam is presented in this paper. The sensitivity of these two novel techniques to micro cracks is discussed and compared to standard traditional sensors. Moreover, the capacity of a DFOS technique to localize cracks and quantify crack openings is also assessed. The results show that the implementation of CWI and DFOS techniques allow the detection of early subtle changes in reinforced concrete structures until crack formation. With their ability to quantify the crack opening, following early detection and localization, DFOS techniques can achieve more effective monitoring of reinforced concrete structures. Contrary to discrete sensors, CWI and DFOS techniques cover larger areas and thus provide more efficient infrastructures asset management and maintenance operations throughout the lifetime of the structure.

6.4.2. Offset Tracking of sensor clock using Kalman filter for wireless network synchronization

Participants: David Pallier, Vincent Le Cam, Qinghua Zhang.

Wireless Sensors Networks (WSN) are more and more used in structural health monitoring applications since they represent a less expensive and non-invasive way to monitor infrastructures. Most of these applications work by merging or comparing data from several sensors located across the structure. These data often comprise measurements of physicals phenomenons evolving with time, such as acceleration and temperature. To merge or compare time-dependent data from different sensors they need to be synchronized so all the samples are time-stamped with the same time reference. An initial synchronization of the sensors is needed because sensors are independent and therefore can not be all started at the same time. Subsequent re-synchronizations are also needed since the sensors keep track of time using their imperfect local clock. This work has been presented in [34].

6.4.3. Wireless implementation of system identification techniques

Participants: Michael Doehler, Mathieu Le Pen, Vincent Le Cam, Laurent Mevel.

Embedded wireless platforms such as the PEGASE platform are appealing and suitable to collect vibration data and then perform off-line and remote computation easily. To obtain detailed modal information of large and very large structures, many sensors would be required to cover the geometry of the structure with a reasonable accuracy. However, when only a limited amount of sensors is available, large structures can be measured in several sensor setups, where some sensors remain fixed and some are moved between different measurement setups. With the sensors connected to different wireless platforms, the synchronous acquisition of data is required. In this paper, a solution of data acquisition synchronization, as well as signal processing for merging the information taking into account the change of sensor positions and environmental variability is presented.

6.4.4. Management of Cloud architectures

Participants: Jean Dumoulin, Laurent Mevel.

Cloud2IR is an autonomous software architecture, allowing multi-sensor connection, dedicated to the long term thermal monitoring of infrastructures. The system has been developed in order to cut down software integration time facilitating the system adaptation to each experiment. First, a generic unit, a data management side able to aggregate any sensor data, type or size, automatically encapsulating them in various generic data format such as hierarchical data format or cloud data such as opengis standard. This whole part is also in charge of the acquisition scenario, the local storage management and the network management. Second, a specialized unit where the sensor specific development fitted to experimental requirements are addressed. The system has been deployed on two test sites for more than one year. It aggregates various sensor data issued from infrared thermal cameras, GPS units, pyranometers, weather stations. The software and some results in outdoor conditions are discussed

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. Collaboration with SNCF on Road circuits

Participants: Vincent Le Cam, Arthur Bouché.

The 2 objectives of the Circuit de Voie project aimed to detect the phenomenon of deshuntage are, with SNCF Innovation Research, develop criteria and models to detect, in real time, the appearance of the phenomenon, and implementing in one of several PEGASE boxes spread over several test sites these models and comparison indicators.

3 criteria have been developed and validated in simulation on real dataset: 1 criterion in residual power on the spectral band of the harmonic of rank 3, a criterion of spectral shape recognition typical in case of bad deshuntage, a statistical criterion on the RMS component of the residual signal. Future work is envisaged in 2020 to go further in comparing these models with real field data and comparison with other detection systems. Several PEGASE units have been built, deployed and implemented for one-off or long-term measurement phases, including during deshuntage tests conducted by SNCF teams.

7.1.2. Collaboration with SNCF Reseau

Participants: Vincent Le Cam, Arthur Bouché.

SNCF has commissioned 5 new DETECTEAU water level sensors adapted to the conditions of nozzles and waterways in the rail network. From a technological point of view the sensor is of small size and very weak consumption. DETECTEAU communicates according to the LORA network. From September to November 2019, one to 3 sites of LGV Paris East will probably be deployed. Scientifically a dynamic sending algorithm has been implemented, taking into account the dynamics of the watercourse (sending more information if there are phases of flood or recession). As it stands, the DETEC-TEAU project is opening the field, probably for 2020, to a more scientific follow-up of the project where the data collected will feed watershed flow models that SNCF wishes to qualify.

7.1.3. Collaboration with SNCF : Hot boxes detection

Participants: Jean Dumoulin, Thibaud Toullier.

The main strategic issue is the maintenance in operational condition of the Hot Box Detectors (DBC). The removal of the DBC from the track is part of Tech4Rail's ambition: reducing equipment to the track. The innovation aimed at in this project is to study and develop a measurement solution to be deployed at the edge of a lane out of danger zone and independent of track equipment. Among the scientific obstacles identified are the following three:

- the behavior of the measurement system in deteriorated meteorological conditions in a real site
- the design and implementation of an automated prototype for in-situ deployment (connection to an existing announcement system, hardware packaging of the system, study and design of a scalable software solution allowing pre-processing data).
- the development of automatic processing tools for the analysis of massive data generated by in-situ measurement systems

7.1.4. Contract with SIEMENS : Poof of Concept monitoring coupled with prediction model for deicing metro lane surface

Participants: Jean Dumoulin, Nicolas Le Touz, Thibaud Toullier.

This proof of concept aims at combining real site monitoring solutions with adjoint state FE thermal model approach to predict optimal heating required to preserve surface from icing in winter conditions. Furthermore, we introduced in our prediction model connection with in-line weather forecast provided by Meteo France Geoservice at different time horizon and spatial scale.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. SYSIFE

Participant: Ivan Guéguen.

Type: CPER + REVES project funding

Objectif: Development of a test bench for railways testing

Duration: 2019 - 2020

Coordinator : IFSTTAR

Partners: Cerema, ColasRail, Edilon, SNCF Réseau, Railenium, Vossloh

Inria contact: Ivan Guéguen

8.1.2. SHM-TGROUT

Participant: Xavier Chapeleau.

Type: Weamec regional cluster

Objectif: to assess the suitability of several non-destructive methods to detect and track the damage for metal pipes.

Duration: 2019 - 2020

Coordinator : IFSTTAR

Partners: University of Nantes, STX

Inria contact: Xavier Chapeleau

Abstract:

The cement bond between metal pipes is a very common technique in the offshore environment, particularly in the "oil and gas" sector. This technique has been used in the offshore wind sector installed to connect the structure (jacket or monopile) to its foundation. A small-scale sample of this type of cement connection was sized, and instrumented it with several technologies. sensors (including fiber optic sensors) and subjected it to axial fatigue stresses. Although the results of the instrumentation are still in operation, the damage could be detected by the various methods tested. A new trial is planned in the first half of 2020 to confirm the results obtained.

8.1.3. MUSIWIND

Participants: Xavier Chapeleau, Laurent Mevel, Frederic Gillot.

Type: RFI WIZE

Objectif: Qualify a very high precision sensor for vibratory monitoring of wind turbines, develop monitoring algorithms using SSI methods and validation indicators

Duration: 12 months in 2020

Coordinator : IFSTTAR

Partners: Inria, SERCELL, VALOREM

Inria contact: Xavier Chapeleau

Abstract: Structural health monitoring of wind turbines is becoming a real economic issue for the managers of these structures. Indeed, they are more and more demanding of new structural health control techniques that enable the implementation of an automated and planned monitoring strategy to ensure the structural integrity of their wind turbines throughout their lifetime, particularly in the case of exceptional events such as storm or earthquake. In this business sector where innovation is crucial to stay competitive, the project MusiWind aims at the hardware, software and scientific development of a new device for monitoring the structural integrity of wind turbines and their qualification in real conditions. Through a multi-sensor approach, the project integrates in particular the newQuietSeisTM low-noise accelerometer (developed by SERCEL) with a generic data acquisition card Pegase 3 (developed by IFSTTAR) on which is embedded innovative signal processing (data analysis) developed by the Ifsttar / Inria I4S joint research team. Statistical inference algorithms meant to extract structural information under ambient excitation. The originality of the project will be to develop identification methods as well as multi-varied damage indicators that merge data froms ensors of different types and qualities, as well as the fusion of complementary physical characteristics.

8.1.4. SURFEOL: SURveillance et Fiabilité des Fondations d'EOLiennes

Participants: Xavier Chapeleau, Michael Doehler, Laurent Mevel, Flavien Bouché.

The regional project SURFEOL was in collaboration with les Chantiers de l'Atlantique and ended in 2017. Many months of data were collected. Three main axes were investigated.

- Study of monitoring of off shore wind turbines
- Laboratory experiments for fatigue monitoring using fiber optic sensors
- Development of a monitoring system based on optical gages and test in real conditions on a marine buoy

A Master 2 internship was dedicated on the analysis of multiple months of data by means of data analysis and subspace identification techniques.

8.1.5. 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 objective is to correct the time drift of the quartz in wireless sensor networks. Quartz modelizations, test platforms under real GPS conditions have been built. First results are based on Kalman algorithms to correct drift[34].

8.1.6. Collaboration with GeM

Participants: Laurent Mevel, Michael Doehler.

I4S' PhD student Md Delwar Hossain Bhuyan has done his PhD on damage localization on civil structures in collaboration with GeM (Institute of Civil and Mechanical Engineering), Université de Nantes, and successfully defended in November 2017. In the follow-up, a mockup of the Saint Nazaire bridge has been funded by GeM in 2018 for damage localization, and tests on it are ongoing [25].

8.1.7. Vibration analysis by video image processing for civil engineering structure monitoring

Participants: Bian Xiong, Qinghua Zhang.

- Type: ARED (Allocations de Recherche Doctorale)
- Objective: to develop video-based methods for civil engineering structure monitoring.
- Duration: 2018 - 2021
- Coordinator : Inria
- Partners: IFSTTAR
- Inria contact: Qinghua Zhang
- Abstract:

  The I4S team develops real-time vibration analysis methods for the monitoring of civil engineering structures (bridges, buildings, etc.), usually based on mechanical sensors integrated into the monitored structures. In parallel, the team works also on image processing techniques for non-destructive testing of civil engineering construction materials. This PhD project, co-supervised with Vincent Baltazart (IFSTTAR researcher), aims to combine the two approaches in order to develop a method of vibration analysis based on image processing. Given a sequence of images of the structure to be monitored, the motion signal of the structure is derived from video image analysis, then methods of vibration analysis are applied to this motion signal. Such a solution will have the advantage of avoiding the integration of mechanical sensors into monitored structures and simplifying the maintenance of the monitoring system

8.2. National Initiatives

8.2.1. CEA List : Acoustic High Frequency synchronous and wireless

Participants: Vincent Le Cam, Arthur Bouché.

In the area of infrastructure, strengthening links with CEA-LIST and Alstom-Rail will focus on non-destructive ultrasonic testing methods for rails. We will focus in particular on the opening of cracks in the passage of the trains, which requires a very precise synchronization of the various sensors. In 2019 the first tests of validation on the site of Bar le Duc with the help of the prototype were conclusive: capacities to emit and receive ultrasonic waves in 1.4 km of rail by perfectly synchronized materials (until the microsecond UT). In 2020 the objectives of the future contract will be:

- make several boxes to carry out more complete tests
- conducting qualification test campaigns (according to CDC Alstom)
- upgrade the high frequency daughter card (with PEGASE 3 more globally)

8.2.2. 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. A smart logger and a protype have been developed and presented. A full autonomous system has been studied and developed for in-situ measurement on existing building envelope. In parallel, thermal resistance estimation method was studied. First experiments were carried out with a first generation prototype in 2019. For this purpose different instrumented building walls were built and qualified at CSTB before carrying out in-situ evaluations of the prototype.

8.3. European Initiatives

8.3.1. FP7 & H2020 Projects

8.3.1.1. 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 Marie-Sklodowska Curie Actions Innovative Training Networks)

Type of Action: MSCA-ITN-ETN

Objective: Improve energy performance of building design

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: This thesis work aims to develop and validate a method for monitoring crack openings using distributed fiber optics strain measurements. First, the various existing theories on strain transfer from the host material to the optical fiber are presented, with their validity domain. The problem of perfect interfacial bonding is then studied and a three-layer analytical model capable of handling imperfect bonding case is proposed. This model is then generalized to multi-layer systems. Experimental studies validating this new model are presented. They show that it is possible to monitor crack openings up to 1 mm with an error of less than 10% for a fiber optic cable glued on the surface. Cables embedded in concrete show less accurate results. The type of cable, the bonding length and the hardening of the concrete material also influence the accuracy of the estimated crack openings. Finally, the results of case studies on laboratory-size reinforced concrete samples are presented. They show the optical fibers capacity to detect cracks as early as ultrasonic sensors and to monitor the opening of multiple micro cracks.

8.3.1.2. DESDEMONA(DEtection of Steel Defects by Enhanced MONitoring and Automated procedure for self-inspection and maintenance)

Participants: Jean Dumoulin, Laurent Mevel, Michael Doehler, Xavier Chapeleau.

Call: H2020 -Call: RFCS-2017 (Call of the research programme of the Research Fund for Coal and Steel - 2017)

Type of Action: RFCS-RPJ (Research project)

Objective: DESDEMONA objective is the development of novel design methods, systems, procedure and technical solution, to integrate sensing and automation technologies for the purpose of self-inspection and self-monitoring of steel structures.

Duration: 36 months since 2018 June 1st

Coordinator: Pr. Vincenzo Gatulli (La Sapienza University of Rome)

Academic and industrial Partners: Sapienza Università di Roma (Italy), Universidad de Castilla – La Mancha, (Spain), Universidade do Porto (Portugal), Università di Pisa (Italy), IFSTTAR (France), Aiviewgroup srl (Italy), Sixense systems (France), Ecisa compania general de construcciones sa (Spain), Università di Cassino e del Lazio Meridionale (Italy), Universidad de Alicante (Spain), Inria (France).

Inria contact: J. Dumoulin and L. Mevel

Website: http://www.desdemonaproject.eu

Abstract: DESDEMONA objective is the development of novel design methods, systems, procedure and technical solution, to integrate sensing and automation technologies for the purpose of self-inspection and self-monitoring of steel structures. The approach will lead to an increment of the service life of existing and new steel civil and industrial infrastructure and to a decrease in the cost associated to inspections, improving human activities performed in difficult conditions, safety and workers' potential by the use of advanced tools. The research aims to expand beyond the current state-of-the-art new high-quality standard and practices for steel structure inspection and maintenance through the interrelated development of the following actions: i) steel structure geometry and condition virtualization through data fusion of image processing, thermography and vibration measurements; ii) developing a procedure for steel defect detection by robotic and automatic systems such as Unmanned Aerial Vehicles (UAV) and ground mobile robots iii) embedding sensor systems to revalorize and transform steel elements and structures into self-diagnostic (smart) elements and materials even through nanotechnologies, iv) realizing an experimental lab-based apparatus and a series of case studies inspected by intelligent and robotic systems. The project outcome will have an impact on the reduction of the cost of steel structures inspection and maintenance and on the increase of user safety and comfort in industrial and civil environment. The proposal with a multidisciplinary approach fulfils the objectives of the Strategic Research Agenda of the European Steel Technology Platform.

8.3.2. Collaborations in European Programs, Except FP7 & H2020

8.3.2.1. 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

Objective: Quantifying the value of structural health monitoring

Duration: 11/2014 - 4/2019

Coordinator: S. Thoens (DTU Denmark)

Partner: 29 countries, see https://www.cost.eu/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.

8.4. International Initiatives

8.4.1. Collaboration with University of British Columbia, Canada

Participants: Laurent Mevel, Michael Doehler, Alexander Mendler.

Alexander Mendler's PhD thesis started in September 2018 co-supervised by M. Doehler and C. Ventura. A. Mendler spent 6 months in Rennes in 2019 thanks to a MITACS grant.

8.4.2. 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 at BAM since 2016. Besides the supervision of the PhD, collaboration on temperature robustness is ongoing with BAM [18], [24].

8.4.3. Collaboration with Technical University of Denmark (DTU)

Participants: Michael Doehler, Laurent Mevel.

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, Aalborg University's PhD student Lijia Long is involved.

8.4.4. Collaboration with Aalborg University, Denmark

Participant: Michael Doehler.

Together with Structural Vibration Solutions, collaboration with Aalborg University (professor Lars Damkilde, Department of Civil Engineering) happened during the PhD of Szymon Gres on damage detection methods, with current conference publications [29], [30]. The PhD has been defended on November 19, 2019.

8.4.5. Collaboration with Laval University, Canada

Participant: Jean Dumoulin.

In the Framework of On Duty Project (http://www.ondutycanada.ca) we are working on Non Destructive Testing techniques and automation of inspection process. Jean Dumoulin spent 10 days in Canada in 2019 devoted to corrosion detection by active infrared thermography NDT approach.

8.5. International Research Visitors

8.5.1. Visits of International Scientists

Szymon Gres visited us for 2 months from January to February 2019 during his thesis.

A. Mendler got a 6 month MITACS grant to visit us from May to October 2019.

8.5.1.1. Research Stays Abroad

J. Dumoulin was with University Laval and with CNR IREA in Fall 2019.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific events selection

9.1.1.1. 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/)
- Co-organizer of sub-Program group PG4: Earth Surface Investigation Method at General Assembly of EGU 2019 (http://www.egu2019.eu/).

Q. Zhang is

- member of the international program commitee (technical associate editor) of the 21st IFAC World Congress that will take place in Berlin, Germany, July 12-17, 2020.
- member of IFAC Technical Committee on Modelling, Identification and Signal Processing (TC 1.1).
- member of IFAC Technical Committee on Adaptive and Learning Systems (TC 1.2).
- member of IFAC Technical Committee on Fault Detection, Supervision and Safety of Technical Processes (TC 6.4).

L. Mevel is

- member of the EWSHM scientific committee.

V. Le Cam is

- member of the IWSHM scientific committee.
- head and general secretary of the EWSHM scientific committee.
- member of the Asian Pacific Workshop

M. Doehler is

- member of IFAC Technical Committee on Modelling, Identification, and Signal Processing.
- member of the IOMAC scientific committee.

9.1.1.2. Reviewers

M. Doehler was reviewer for ECC 2020.

J. Dumoulin was reviewer for EGU 2019, QIRT ASIA 2019 and SFT 2019.

Q. Zhang was reviewer for CDC 2019, IFAC Word Congress 2020.

9.1.2. Journal

9.1.2.1. 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 Executive Editor for the journal Geoscientific Instrumentation and Data Systems.

9.1.2.2. Reviewer - Reviewing Activities

L. Mevel was reviewer for Mechanical Systems and Signal Processing, Engineering Structures.Structural Control and Health Monitoring

M. Doehler was reviewer for Mechanical Systems and Signal Processing, Engineering Structures, Philosophical Transactions A Mathematical, Physical and Engineering Sciences, Computers and Structures, Structural Health Monitoring.

J. Dumoulin was reviewer for Quantitative Infrared Thermography Journal, GI Journal (EGU), SFT conference, Engineering Geology, International Journal of Pavement Research and Technology, Remote sensing of environment, Structural Health Monitoring, Composites Structures

9.1.3. Invited Talks

L. Ibos and J. Dumoulin, « Thermographie infrarouge: du laboratoire vers les mesures de terrain », Atelier Métrologie Thermique Avancée, SFT 2019 (Congrès annuel de la Société Française de Thermique), Nantes, France, 3 – 6 juin 2019.

9.1.4. Leadership within the Scientific Community

V. Le Cam organized the 2nd SHM France meeting in Nantes, together with CEA and Precend.

9.1.5. Scientific Expertise

V. Le Cam did an expertise in the context of EFFICACITY IRT. The expertise was about the instrumentation of RATP station and the control of the speed of escalators.

V. Le Cam helped SNCF Reseau to disseminate at industrial results the monitoring of axle counters developed in 2018.

Arthur Bouché did an expertise for SNCF about electromagnetism and axle counters.

Ivan Guéguen did an expertise for ASCQUER for labelization of Lacroix Millenium lights for construction sites.

9.1.6. 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.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

J. Dumoulin

- Licence Professionnelle TAM (Techniques Avancées en Maintenance): thermographie infrarouge active, 24h, Université Paris-Est Créteil (UPEC), 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.
- With Nicolas Le Touz lecture on Solar Hybrid Roads at SMARTIE ETN Training week 2019
- Lecture on Long term thermal monitoring of Structures at SMARTIE ETN Training week 2019
- Lecture on IR for inspection and/or thermal monitoring of infrastructures: scope of application, technical solutions and analysis methods, at QIRT ASIA 2019 courses day.
- Lecture on Active Thermography for NDE at APESS 2019 (Asia-Pacific-Euro Summer School on Smart Structures).

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
- Polytech la Roche sur Yon, 8h embeded wireless algorithms §8h for David Pallier

M. Doehler

- Cycle préparatoire intégré, STPI, mathématiques, 48h TD, INSA Rennes, France

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

9.2.2. Supervision

PhD : Thibaud 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,defended on 6th November 2019. [15].

PhD : Antoine Bassil, *Fibre-optic sensor for fatigue monitoring*, D. Leduc, O. Abraham and X. Chapeleau. Ecole doctorale SPIGA, Université de Nantes, defended on 26 November 2019.

PhD : Bian Xong, *Vibration analysis by video image processing for civil engineering structure monitoring*, Q. Zhang, V. Balthazar. Ecole doctorale MathsTIC, Université de Rennes 1, since October 2018.

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 : David Pallier, *Sensor Enhancement to Augmented Usage and Reliability*, S. Pillement, IETR, V. Le Cam, Ecole doctorale MathSTIC

PhD : Alexander Mendler, *Vibration-based structural health monitoring of road bridges*, C. Ventura and M. Doehler. University of British Columbia, Vancouver, Canada, since September 2018.

Ludovic Gavérina post-doctoral project on in-situ measurement of thermal resistance of building envelopes, J. Dumoulin, march 2017- february 2019.

Nicolas Le Touz post-doctoral project on in-situ measurement of thermal resistance of building envelopes, J. Dumoulin, December 2018- august 2019.

9.2.3. Juries

Q. Zhang participated in the following PhD defense committees:

- Missie Aguado Rojas at CentraleSupélec, on June 14, 2019.
- Tzila Ajamian at Centrale Nantes on October 24, 2019.

9.3. Popularization

9.3.1. Internal or external Inria responsibilities

L. Mevel is member of CLHSCT committee in Rennes.

L. Mevel is member of Comité de centre committee in Rennes.

L. Mevel is Head of Science in 2nd of Inria Rennes Center.

V. Le Cam is Head of SII LAb at IFSTTAR Nantes.

J. Dumoulin is scientific Head SII LAb at IFSTTAR Nantes.

9.3.2. Articles and contents

9.3.2.1. Encycopedia article on nonlinear system identification

In the second edition of Encyclopedia of Systems and Control, published by Springer in September 2019, the article on nonlinear system identification authored by Q. Zhang has been updated with the latest progresses on this topic.

Nonlinear mathematical models are essential tools in various engineering and scientific domains, where more and more data are recorded by electronic devices. How to build nonlinear mathematical models essentially based on experimental data is the topic of this entry. Due to the large extent of the topic, this entry provides only a rough overview of some well-known results, from gray-box to black-box system identification. DOI: https://doi.org/10.1007/978-1-4471-5102-9_104-2

9.3.3. Interventions

M. Doehler and Q. Zhang participated the outreach activity "J'peux pas, j'ai informatique" for school classes on April 2, 2019.

9.3.4. Creation of media or tools for science outreach

9.3.4.1. Showroom demonstrations

A damage localization mockup has been developed and installed in the showroom of Inria Rennes. It has been the support for a demonstration for the outreach activity "J'peux pas, j'ai informatique" for school classes on April 2, 2019, showing that computer science also can be related to physics and statistics. An ADT Carnot has been funded on that topic and will help the maturation of such project.

10. Bibliography

Major publications by the team in recent years

[1] 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^{o} 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^{o} 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^{o} 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^{o} 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^{o} 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^{o} 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^o 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^o 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^o 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^o 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^o 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^o 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^o 1, p. 141 - 148, https://hal.inria.fr/inria-00365991

Publications of the year

Doctoral Dissertations and Habilitation Theses

[15] T. TOULLIER.*Caractérisation conjointe de la température et des propriétés radiatives des objets par thermographie infrarouge multispectrale*, Université Rennes 1, November 2019, https://tel.archives-ouvertes.fr/tel-02389051

Articles in International Peer-Reviewed Journal

[16] S. ALLAHDADIAN, M. DÖHLER, C. VENTURA, L. MEVEL.*Towards robust statistical damage localization via model-based sensitivity clustering*, in "Mechanical Systems and Signal Processing", 2019, vol. 134, p. 1-25 [*DOI :* 10.1016/J.YMSSP.2019.106341], https://hal.inria.fr/hal-02293021

[17] A. BASSIL, X. WANG, X. CHAPELEAU, E. NIEDERLEITHINGER, O. ABRAHAM, D. LEDUC.*Distributed fiber optics sensing and coda wave interferometry techniques for damage monitoring in concrete structures*, in "Sensors", January 2019, vol. 19, n^o 2, p. 1-15 [*DOI :* 10.3390/S19020356], https://hal.archives-ouvertes.fr/hal-02007049

[18] M. D. H. BHUYAN, G. GAUTIER, N. LE TOUZ, M. DÖHLER, F. HILLE, J. DUMOULIN, L. MEVEL.*Vibration-based damage localization with load vectors under temperature changes*, in "Structural Control and Health Monitoring", September 2019, vol. 26, n^o 11, e2439 [*DOI :* 10.1002/STC.2439], https://hal.inria.fr/hal-02293057

[19] A. CRINIÈRE, J. DUMOULIN, L. MEVEL.*Management of local multi-sensors applied to SHM and long term infrared monitoring: Cloud2IR implementation*, in " Quantitative InfraRed Thermography Journal", March 2019, vol. 16, n^{o} 1, p. 55-73 [*DOI :* 10.1080/17686733.2018.1519752], https://hal.inria.fr/hal-02293809

[20] C. IBARRA-CASTANEDO, M. KLEIN, M. LAVOIE, D. PROTEAU, J. DUMOULIN.*Evaluation of Impact of Hot-Mix Asphalt Density Differentials on Thermal Streak Phenomenon by Passive Infrared Thermography*, in "Journal of Materials in Civil Engineering", October 2019, vol. 31, n^{o} 10, 04019215 [*DOI :* 10.1061/(ASCE)MT.1943-5533.0002822], https://hal.inria.fr/hal-02294005

[21] Y. YANG, T. V. WU, A. SEMPEY, J. DUMOULIN, J. C. BATSALE.*Short time non-destructive evaluation of thermal performances of building walls by studying transient heat transfer*, in "Energy and Buildings", February 2019, vol. 184, p. 141-151 [*DOI :* 10.1016/J.ENBUILD.2018.12.002], https://hal.archives-ouvertes.fr/hal-02044900

[22] Q. ZHANG, L. LJUNG, R. PINTELON.*On Local LTI Model Coherence for LPV Interpolation*, in "IEEE Transactions on Automatic Control", October 2019, 6 [*DOI :* 10.1109/TAC.2019.2948898], https://hal.inria.fr/hal-01758337

[23] Q. ZHANG, L. ZHANG.*Stability Analysis of the Kalman Predictor*, in "International Journal of Control", July 2019, p. 1-16 [*DOI :* 10.1080/00207179.2019.1638971], https://hal.inria.fr/hal-02373906

Invited Conferences

[24] M. BAESSLER, M. D. H. BHUYAN, F. HILLE, E. VIEFHUES, M. DÖHLER, L. MEVEL.*Impact of environmental based effects on SHM strategies*, in "SEMC 2019 - 7th International Conference on Structural Engineering, Mechanics and Computation", Cape Town, South Africa, September 2019, p. 1-6, https://hal.inria.fr/hal-02290909

International Conferences with Proceedings

[25] M. D. H. BHUYAN, M. DÖHLER, Y. LECIEUX, C. LUPI, J.-C. THOMAS, F. SCHOEFS, F. HILLE, L. MEVEL.*Statistical subspace based damage localization on Saint-Nazaire Bridge mock-up*, in "IOMAC 2019 - 8th International Operational Modal Analysis Conference", Copenhagen, Denmark, May 2019, p. 1-9, https://hal.inria.fr/hal-02143530

[26] M. D. H. BHUYAN, N. LE TOUZ, G. GAUTIER, M. DÖHLER, F. HILLE, J. DUMOULIN, L. MEVEL.*Load vector based damage localization with rejection of the temperature effect*, in "IOMAC 2019 - 8th International Operational Modal Analysis Conference", Copenhagen, Denmark, May 2019, p. 1-10, https://hal.inria.fr/hal-02143742

[27] J. DUMOULIN.*Uncooled infrared thermal camera for thermal monitoring or Non-Destructive Testing of Civil Engineering structures*, in "ANCRiSST 2019 - 14th International Workshop on Advanced Smart Materials and Smart Structures Technology", Rome, Italy, July 2019, https://hal.inria.fr/hal-02294079

[28] L. GAVÉRINA, T. HA, J. WAEYTENS, V. FEUILLET, J.-L. MANCEAU, L. PFEIFFER, J.-P. MONCHEAU, M. MARCHETTI, L. IBOS, J. DUMOULIN.*Study and designed of an active infrared system for in-situ characterization of thermal resistance of building envelopes*, in "QIRT ASIA 2019 - 3rd Asian Conference on Quantitative InfraRed Thermography", Tokyo, Japan, July 2019, https://hal.inria.fr/hal-02294671

[29] S. GRES, M. DÖHLER, P. ANDERSEN, L. DAMKILDE, L. MEVEL.*Hankel matrix normalization for robust damage detection*, in "IOMAC 2019 - 8th International Operational Modal Analysis Conference", Copenhagen, Denmark, May 2019, p. 1-8, https://hal.inria.fr/hal-02143749

[30] S. GRES, M. DÖHLER, P. ANDERSEN, L. MEVEL.*Variance computation of MAC and MPC for real-valued mode shapes from the stabilization diagram*, in "IOMAC 2019 - 8th International Operational Modal Analysis Conference", Copenhagen, Denmark, May 2019, p. 1-9, https://hal.inria.fr/hal-02143765

[31] M. LE PEN, A. BOUCHÉ, I. GUÉGUEN, M. DÖHLER, L. MEVEL, V. LE CAM.*Phased and synchronous sampling between multiple smart network sensors for modal assessment of large structures*, in "IWSHM 2019 - 12th International Workshop on Structural Health Monitoring", Stanford, CA, United States, September 2019, p. 1-9, https://hal.inria.fr/hal-02290901

[32] N. LE TOUZ, J. DUMOULIN.*Study of an optimal command law combining weather forecast and energy reduction for transport structure surface de-icing by Joule effect*, in "ANCRiSST 2019 - 14th International Workshop on Advanced Smart Materials and Smart Structures Technology", Rome, Italy, July 2019, https://hal.inria.fr/hal-02294081

[33] A. MENDLER, S. ALLAHDADIAN, M. DÖHLER, L. MEVEL, C. VENTURA.*Minimum detectable damage for stochastic subspace-based methods*, in "IOMAC 2019 - 8th International Operational Modal Analysis Conference", Copenhagen, Denmark, May 2019, p. 1-11, https://hal.inria.fr/hal-02142994

[34] D. PALLIER, V. LE CAM, A. BOUCHE, S. PILLEMENT, Q. ZHANG, L. MEVEL.*Offset Tracking of sensor clock using Kalman filter for wireless network synchronization*, in "IWSHM 2019 - 12th International Workshop on Structural Health Monitoring", Stanford, California, United States, IWSHM 2019, September 2019, paper #516, https://hal.archives-ouvertes.fr/hal-02155889

[35] A. RUAS, F. BOURQUIN, J. DUMOULIN, B. LEBENTAL.*Sense-City innovation lab supporting recent advances on Monitoring of urban operations*, in "EGU 2019 - European Geoscience Union", Vienne, Austria, April 2019, https://hal.inria.fr/hal-02294567

[36] T. TOULLIER, J. DUMOULIN, L. MEVEL.*Study of complementary multi-sensors data influence on infrared thermography measurements for in-situ long-term monitoring*, in "2019 - Multimodal Sensing and Artificial Intelligence: Technologies and Applications", Munich, Germany, SPIE, June 2019, vol. 11059, p. 1-10 [*DOI :* 10.1117/12.2526229], https://hal.archives-ouvertes.fr/hal-02264743

[37] E. VIEFHUES, M. DÖHLER, Q. ZHANG, F. HILLE, L. MEVEL.*Subspace-based damage detection with rejection of the temperature effect and uncertainty in the reference*, in "IOMAC 2019 - 8th International Operational Modal Analysis Conference", Copenhagen, Denmark, May 2019, p. 1-11, https://hal.inria.fr/hal-02143745

Conferences without Proceedings

[38] T. TOULLIER, J. DUMOULIN, V. BOURGEOIS.*Comparative study of moving train hot boxes pre-detection and axles counting by in-situ implementation of two infrared cameras*, in "QIRT ASIA 2019 - Quantitative InfraRed Thermography Conference", Tokyo, Japan, July 2019, https://hal.archives-ouvertes.fr/hal-02264672

[39] T. TOULLIER, J. DUMOULIN, L. MEVEL.*Étude et développement d'un simulateur d'échanges radiatifs dans des scènes 3D statiques et dynamiques surveillées par thermographie infrarouge*, in "SFT 2019 -

27eme congrès français de thermique", Nantes, France, June 2019, p. 1-8, https://hal.archives-ouvertes.fr/hal-02264749

[40] Q. ZHANG, F. GIRI, T. AHMED-ALI.*Regularized Adaptive Observer to Address Deficient Excitation*, in "13th IFAC Workshop on Adaptive and Learning Control Systems (ALCOS 2019)", Winchester, United Kingdom, December 2019, https://hal.archives-ouvertes.fr/hal-02414207

Other Publications

[41] J. DUMOULIN, L. GAVÉRINA, C. IBARRA-CASTANEDO, M. KLEIN, X. MALDAGUE.*Non Destructive Testing of CFRP by active infrared thermography using uncooled IRFPA camera mounted on mobile system*, April 2019, EGU 2019 - European Geoscience Union, Poster, https://hal.inria.fr/hal-02294153

[42] M. RIBAUD, C. BLANCHET-SCALLIET, F. GILLOT, C. HELBERT.*Robustness kriging-based optimization*, February 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01829889

[43] T. TOULLIER, J. DUMOULIN, L. MEVEL.*Sensitivity of different methods for simultaneous evaluation of emissivity and temperature through multispectral infrared thermography simulation*, April 2019, vol. 21, 1, EGU 2019 - European Geoscience Union, Poster, https://hal.archives-ouvertes.fr/hal-02264677

References in notes

[44] 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^{o} 2, p. 189–202, http://dx.doi.org/10.1016/S1367-5788(02)00029-9

[45] B. DELYON, A. JUDITSKY, A. BENVENISTE.*On the relationship between identification and local tests*, IRISA, May 1997, n^{o} 1104, ftp://ftp.irisa.fr/techreports/1997/PI-1104.ps.gz

[46] M. JAULENT.*The inverse scattering problem for LCRG transmission lines*, in "Journal of Mathematical Physics", December 1982, vol. 23, n^{o} 12, p. 2286-2290

[47] G. L. LAMB.*Elements of Soliton Theory*, John Wiley & Sons, New York, 1980

[48] M. OUMRI.*Fault diagnosis of wired electric networks by reflectometry*, Université Paris Sud - Paris XI, May 2014, https://tel.archives-ouvertes.fr/tel-01165039

[49] C. R. PAUL.*Analysis of multiconductor transmission lines*, Wiley, New York, 2008

[50] 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^{o} 10, p. 3730-3737, http://dx.doi.org/10.1109/TAP.2011.2163772

[51] P. VAN OVERSCHEE, B. DE MOOR.*Subspace Identification for Linear Systems*, Kluwer Academic Publishers, Boston, 1996

[52] 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 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

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
A1.1.9. - Fault tolerant systems
A1.3. - Distributed Systems
A1.3.5. - Cloud
A1.3.6. - Fog, Edge
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
A9.7. - AI algorithmics

Other Research Topics and Application Domains:

B3.2. - Climate and meteorology
B3.3.1. - Earth and subsoil
B8.2. - Connected city
B9.5.6. - Data science

1. Team, Visitors, External Collaborators

Research Scientist

Gabriel Antoniu [Team leader, Inria, Senior Researcher, HDR]

Faculty Members

Luc Bougé [École normale supérieure de Rennes, Professor, HDR]
Alexandru Costan [INSA Rennes, Associate Professor, HDR]

Technical Staff

Ovidiu-Cristian Marcu [Inria, Engineer, until Oct 2019]

PhD Students

Nathanaël Cheriere [École normale supérieure de Rennes, until Nov 2019]
Paul Le Noac'h [INSA Rennes, until Feb 2019]
Daniel Rosendo [Inria, from Oct 2019]

Post-Doctoral Fellow

Pedro de Souza Bento Da Silva [INSA Rennes]

Visiting Scientists

José Aguilar Canepa [Instituto Politécnico Nacional, from Sep 2019 until Nov 2019]

Mario Rivero Angeles [Instituto Politécnico Nacional, from Jun 2019 until Jul 2019]
Edgar Romo Montiel [Instituto Politécnico Nacional, from Sep 2019 until Nov 2019]

Administrative Assistant

Gaëlle Tworkowski [Inria]

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 *scalable 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 [26], [23] and Exascale HPC [25] research communities. The Big Data area, 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 [30], 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 center of Rennes [29]: *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 [35] parallel programming paradigm. It is currently used on most cloud platforms, following the trend set up by Amazon [22]. 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 large-scale simulations running on post-Petascale supercomputers. Such applications exhibit requirements clearly identified by international panels of experts like IESP [28], EESI [24], ETP4HPC [25]. 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 [31]). 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 HPC and Big Data

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 ecosystems is crucial to the future.

Our objective is premised on the idea that we must explore the ways in which the major challenges associated with Big Data analytics intersect with, impact, and potentially change the directions now in progress for achieving Exascale computing.

In particular, a key milestone will be to achieve convergence through common abstractions and techniques for data storage and processing in support of complex workflows combining simulations and analytics. Such application workflows will need such a convergence to run on hybrid infrastructures combining HPC systems and clouds (potentially in extension to edge devices, in a complete digital continuum).

Collaboration. *This axis is addressed in close collaboration with María Pérez (UPM), Rob Ross (ANL), Toni Cortes (BSC), Several groups at Argonne National Laboratory and NCSA (Franck Cappello, Rob Ross, Bill Kramer, Tom Peterka).*

Relevant groups with similar interests are the following ones.

- *The group of Jack Dongarra, Innovative Computing Laboratory at University of Tennessee, who is leading international efforts for the convergence of Exascale Computing and Big Data.*
- *The group of Satoshi Matsuoka, RIKEN, working on system software for clouds and HPC.*
- *The group of Ian Foster, Argonne National Laboratory, working on on-demand data analytics and storage for extreme-scale simulations and experiments.*

3.1.1. High-performance storage for concurrent Big Data applications

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.

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, multiblob 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. Towards unified data processing techniques for Extreme Computing and Big Data applications

In the high-performance computing area (HPC), the need to get fast and relevant insights from massive amounts of data generated by extreme-scale computations led to the emergence of *in situ processing*. It allows data to be visualized and processed in real-time on the supercomputer generating them, in an interactive way, as they are produced, as opposed to the traditional approach consisting of transferring data off-site after the end of the computation, for offline analysis. As such processing runs on the same resources executing the simulation, if it consumes too many resources, there is a risk to "disturb" the simulation.

Consequently, an alternative approach was proposed (*in transit processing*), as a means to reduce this impact: data are are transferred to some temporary processing resources (with high memory and processing capacities). After this real-time processing, they are moved to persistent storage.

In the Big Data area, the search for real-time, fast analysis was materialized through a different approach: stream-based processing. Such an approach is based on a different abstraction for data, that are seen as a dynamic flow of items to be processed. Stream-based processing and in situ/in transit processing have been developed separately and implemented in different tools in the BDA and HPC areas respectively.

A major challenge from the perspective of the HPC-BDA convergence is their joint use in a unified data processing architecture. This is one of the future research challenges that I plan to address in the near future, by combining ongoing approaches currently active in my team: Damaris and KerA. We started preliminary work within the *"Frameworks"* work package of the HPC-Big Data IPL. Further exploring this convergence is a core direction of our current efforts to build collaborative European projects.

3.2. Research axis 2: Cloud and Edge processing

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 storage and processing techniques to cope with the volatile requirements of data-intensive applications on large-scale dynamic clouds in a cost-efficient way, with a particular focus on streaming. More recently, the strong emergence of edge/fog-based infrastructures leads to to additional challenges for new scenarios involving hybrid cloud/fog/edge systems.

Collaboration. *This axis is addressed in close collaboration with María Pérez (UPM), Kate Keahey. (ANL)*

Relevant groups with similar interests include the following ones.

- *The group of Geoffrey Fox, Indiana University, working on data analytics, cloud data processing, stream processing.*
- *The group at RISE Lab, UC Berkeley, working on real-time stream-based processing and analytics.*
- *The group of Ewa Deelman, USC Information Sciences Institute, working on resource management for workflows in clouds.*

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 above-mentioned requirements, several systems have been introduced for stream data processing such as Flink [27], Spark [32], Storm [33], and Google MillWheel [34]. 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.

3.2.2. Efficient Edge, Cloud and hybrid Edge/Cloud data processing

Today, we are approaching an important technological milestone: applications are generating huge amounts of data and are demanding low-latency responses to their requests. Mobile computing and Internet of Things (IoT) applications are good illustrations of such scenarios. Using only Cloud computing for such scenarios is challenging. Firstly, Cloud resources are most of the time accessed through Internet, hence, data are sent across high-latency wide area networks, which may degrade the performance of applications. Secondly, it may be impossible to send data to the Cloud due to data regulations, national security laws or simply because an Internet connection is not available. Finally, data transmission costs (e.g., Cloud provider fees, carrier costs) could make a business solution impractical.

Edge computing is a new paradigm which aims to address some of these issues. The key idea is to leverage computing and storage resources at the "edge" of the network, i.e., on processing units located close to the data sources. This allows applications to outsource task execution from the main (Cloud) processing data centers to the edge. The development of Edge computing was accelerated by the recent emergence of stream processing, a new model for handling continuous flows of data in real-time, as opposed to batch processing, which typically processes bounded datasets offline.

However, Edge computing is not a silver bullet. Besides being a new concept not fully established in the community, issues like node volatility, limited processing power, high latency between nodes, fault tolerance and data degradation may impact applications depending on the characteristics of the infrastructure.

Some relevant research questions are: How much can one improve (or degrade) the performance of an application by performing data processing closer to the data sources rather than performing it in the cloud? How to progress towards a seamless scheduling and execution of a data analytics workflow and break the limitation the current dual approaches used in preliminary efforts in this area, that rely on manual and empirical deployment of the corresponding dataflow operator graphs, using separate analytics engines for centralized clouds and for edge systems respectively?

Our objective is to try to answer precisely such questions. We are interested in understanding the conditions that enable the usage of Edge or Cloud computing to reduce the time to results and the associated costs. While some state-of-the-art approaches advocate either "100% Cloud" or "100% Edge" solutions, the relative efficiency of a method over the other may vary. Intuitively, it depends on many parameters, including network technology, hardware characteristics, volume of data or computing power, processing framework configuration and application requirements, to cite a few. We plan to study their impact on the overall application performance.

3.3. Research axis 3: Supporting AI across the digital continuum

Integrating and processing high-frequency data streams from multiple sensors scattered over a large territory in a timely manner requires high-performance computing techniques and equipments. For instance, a machine learning earthquake detection solution has to be designed jointly with experts in distributed computing and cyber-infrastructure to enable real-time alerts. Because of the large number of sensors and their high sampling rate, a traditional centralized approach which transfers all data to a single point may be impractical. Our goal is to investigate innovative solutions for the design of efficient data processing infrastructures for a distributed machine learning-based approach.

In particular, building on our previous results in the area of efficient stream processing systems, we aim to explore approaches for unified data storage, processing and machine-learning based analytics across the whole digital continuum (i.e., for highly distributed applications deployed on hybrid edge/cloud/HPC infrastructures). Our ZettaFlow project is targeting a startup creation precisely this area.

Collaboration. *This recently started axis is worked out in close collaboration with the group of Manish Parashar, Rutgers University, and with the LACODAM team at Inria, focused on large-scale collaborative data mining.*

4. Application Domains

4.1. Application Domains

The KerData team investigates the design and implementation of architectures for data storage and processing across clouds, HPC and edge-based systems, which address the needs of a large spectrum of applications. The use cases we target to validate our research results come from the following domains.

- Climate and meteorology
- Earth science
- Energy and sustainable development
- Smart cities
- Data science

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Contributions to the ETP4HPC agenda

The KerData team contributed to the new ETP4HPC Strategic Agenda (to appear). It will serve as a reference for the future EU funding strategy for HPC. Gabriel Antoniu served as a co-leader of the Programming Environment working group. He also served as a co-leader of 2 transversal (*"cross-working group"*) research clusters: *"HPC and the Digital Continuum"* and *"Data Everywhere"*. Alexandru Costan served as a member of these groups.

5.1.2. Paper co-authored with the LACODAM team published in a major AI conference

In 2019, Pedro Silva initiated a multi-disciplinary collaboration with the LACODAM Inria team and the team of Manish Parashar at Rutgers University. It addresses Machine Learning in the context of Edge stream processing. The target application is early earthquake detection from motion sensors distributed on the ground.

This collaboration resulted in a co-authored paper titled *Distributed Multi-Sensor Machine Learning Approach to Earthquake Early Warning* [21]. It will be presented at the 34th AAAI Conference on Artificial Intelligence (AAAI-20), a top conference for Machine Learning (CORE Rank: A*). It is the first paper published by the team in a major AI venue.

5.1.3. Awards

Pierre Matri, earned a PhD in May 2018 co-advised by Maria Pérez (Universidad Politécnica de Madrid, UPM), Alexandru Costan (INSA Rennes) and Gabriel Antoniu (Inria). This PhD was defended at UPM and it received the Outstanding PhD Award (*Premio Extraordinario*) of UPM.

6. New Software and Platforms

6.1. Damaris

KEYWORDS: Visualization - I/O - HPC - Exascale - High performance computing

SCIENTIFIC DESCRIPTION: Damaris is a middleware for I/O and data management targeting large-scale, MPI-based HPC simulations. It initially proposed to dedicate cores for asynchronous I/O in multicore nodes of recent HPC platforms, with an emphasis on ease of integration in existing simulations, efficient resource usage (with the use of shared memory) and simplicity of extension through plug-ins. Over the years, Damaris has evolved into a more elaborate system, providing the possibility to use dedicated cores or dedicated nodes to in situ data processing and visualization. It proposes a seamless connection to the VisIt visualization framework to enable in situ visualization with minimum impact on run time. Damaris provides an extremely simple API and can be easily integrated into the existing large-scale simulations.

Damaris was at the core of the PhD thesis of Matthieu Dorier, who received an Accessit to the Gilles Kahn Ph.D. Thesis Award of the SIF and the Academy of Science in 2015. Developed in the framework of our collaboration with the JLESC – Joint Laboratory for Extreme-Scale Computing, Damaris was the first software resulted from this joint lab validated in 2011 for integration to the Blue Waters supercomputer project. It scaled up to 16,000 cores on Oak Ridge's leadership supercomputer Titan (first in the Top500 supercomputer list in 2013) before being validated on other top supercomputers. Active development is currently continuing within the KerData team at Inria, where it is at the center of several collaborations with industry as well as with national and international academic partners.

FUNCTIONAL DESCRIPTION: Damaris is a middleware for data management and in-situ visualization targeting large-scale HPC simulations: - In situ data analysis by some dedicated cores/nodes of the simulation platform - Asynchronous and fast data transfer from HPC simulations to Damaris - Semantic-aware dataset processing through Damaris plug-ins - Writing aggregated data (by hdf5 format) or visualizating them either by VisIt or ParaView

- Participants: Gabriel Antoniu, Lokman Rahmani, Luc Bougé, Matthieu Dorier, Orçun Yildiz and Hadi Salimi
- Partner: ENS Rennes
- Contact: Matthieu Dorier
- URL: https://project.inria.fr/damaris/

6.2. 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 geo-replication, 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

6.3. Pufferbench

KEYWORDS: Distributed Storage Systems - Elasticity - Benchmarking

SCIENTIFIC DESCRIPTION: Pufferbench is a benchmark for evaluating how fast one can scale up and down a distributed storage system on a given infrastructure and, thereby, how viably can one implement storage malleability on it. Besides, it can serve to quickly prototype and evaluate mechanisms for malleability in existing distributed storage systems.

FUNCTIONAL DESCRIPTION: Pufferbench is a benchmark to designed to evaluate whether to use malleable distributed storage systems on a given platform. - It measures the duration of commission and decommission operations. - Its modularity allows to quickly change and adapt each component to the needs of the user. - It can serve as a baseline when implementing commission and decommission mechanisms in a distributed storage system.

RELEASE FUNCTIONAL DESCRIPTION: This is the first release of Pufferbench.

It includes default components for each of the customisable components: - storage: in memory, on drive with file system cache, and on drive without file system cache - network: MPI network - IODispatcher: basic, and with acknowledgements - DataTransferScheduler: basic - DataDistributionGenerator: uniform, and random - MetadataGenerator: Files of same size The diversity of available components enables Pufferbench to fit to multiple use cases.

- Participants: Nathanaël Cheriere, Matthieu Dorier and Gabriel Antoniu
- Partner: ENS Rennes
- Contact: Nathanaël Cheriere
- Publication: hal-01886351
- URL: https://gitlab.inria.fr/Puffertools/Pufferbench/wikis/home

6.4. Tyr

KEYWORDS: Cloud storage - Distributed Storage Systems - Big data

FUNCTIONAL DESCRIPTION: Tyr is the first blob storage system to provide built-in, multiblob transactions, while retaining sequential consistency and high throughput under heavy access concurrency. Tyr offers fine-grained random write access to data and in-place atomic operations.

- Partner: Universidad Politécnica de Madrid
- Contact: Gabriel Antoniu

6.5. Planner

KEYWORDS: Edge elements - Cloud computing - Scheduling

FUNCTIONAL DESCRIPTION: Planner is a middleware for uniform and transparent stream processing across Edge and Cloud. Planner automatically selects which parts of the executiongraph will be executed at the Edge in order to minimize thenetwork cost.

- Partner: ENS Cachan
- Contact: Gabriel Antoniu
- URL: https://team.inria.fr/kerdata/

6.6. KerA

KerAnalytics

KEYWORD: Distributed Storage Systems

FUNCTIONAL DESCRIPTION: A unified architecture for stream ingestion and storage which can lead to the optimization of the processing of Big Data applications. This approach minimizes data movement within the analytics architecture, finally leading to better utilized resources.

- Contact: Gabriel Antoniu

6.7. TailWind

KEYWORDS: Fault-tolerance - Data management. - Distributed Data Management

FUNCTIONAL DESCRIPTION: Replication is essential for fault-tolerance. However, in in-memory systems, it is a source of high overhead. Remote direct memory access (RDMA) is attractive to create redundant copies of data, since it is low-latency and has no CPU overhead at the target. However, existing approaches still result in redundant data copying and active receivers. To ensure atomic data transfers, receivers check and apply only fully received messages. Tailwind is a zero-copy recovery-log replication protocol for scale-out in-memory databases. Tailwind is the first replication protocol that eliminates *all* CPU-driven data copying and fully bypasses target server CPUs, thus leaving backups idle. Tailwind ensures all writes are atomic by leveraging a protocol that detects incomplete RDMA transfers. Tailwind substantially improves replication throughput and response latency compared with conventional RPC-based replication. In symmetric systems where servers both serve requests and act as replicas, Tailwind also improves normal-case throughput by freeing server CPU resources for request processing. We implemented and evaluated Tailwind on RAMCloud, a low-latency in-memory storage system. Experiments show Tailwind improves RAMCloud's normal-case request processing throughput by $1.7\times$. It also cuts down writes median and 99^{th} percentile latencies by 2x and 3x respectively.

- Contact: Gabriel Antoniu

7. New Results

7.1. Convergence HPC and Big Data

7.1.1. Convergence at the data-processing level

Participants: Gabriel Antoniu, Alexandru Costan, Daniel Rosendo.

Traditional data-driven analytics relies on Big Data processing techniques, consisting of batch processing and real-time (stream) processing, potentially combined in a so-called *Lambda architecture*. This architecture attempts to balance latency, throughput, and fault-tolerance by using batch processing to provide comprehensive and accurate views of batch data, while simultaneously using real-time stream processing to provide views of online data.

On the other side, simulation-driven analytics is based on computational (usually physics-based) simulations of complex phenomena, which often leverage HPC infrastructures. The need to get fast and relevant insights from massive amounts of data generated by extreme-scale simulations led to the emergence of in situ and in transit processing approaches: they allow data to be visualized and processed interactively in real-time as data are produced, while the simulation is running.

To support hybrid analytics and continuous model improvement, we propose to combine the above data processing techniques in what we will call the *Sigma architecture*, a HPC-inspired extension of the Lambda architecture for Big Data processing [17]. Its instantiation in specific application settings depends of course of the specific application requirements and of the constraints that may be induced by the underlying infrastructure. Its main conceptual strength consists in the ability to leverage in a unified, consistent framework, data processing techniques that became reference in HPC in the Big Data communities respectively, without however being combined so far for joint usage in converged environments.

The given framework will integrate previously-validated approaches developed in our team, such as Damaris, a middleware system for efficient I/O management and large-scale in situ data processing, and KerA, a unified system for data flow ingestion and storage. The overall objective is to enable the usage of a large spectrum of Big Data analytics and Intelligence techniques at extreme scales in the Cloud and Edge, to support continuous intelligence (from streaming and historical data) and precise insights/predictions in real-time and fast decision making.

7.1.2. Pufferscale: Elastic storage to support dynamic hybrid workflows systems

Participants: Nathanaël Cheriere, Gabriel Antoniu.

User-space HPC data services are emerging as an appealing alternative to traditional parallel file systems, because of their ability to be tailored to application needs while eliminating unnecessary overheads incurred by POSIX compliance. Such services may need to be rescaled up and down to adapt to changing workloads, in order to optimize resource usage. This can be useful, for instance, to better support complex workflows that mix on-demand simulations and data analytics.

We formalized the operation of rescaling a distributed storage system as a multi objective optimization problem considering three criteria: load balance, data balance, and duration of the rescaling operation. We proposed a heuristic for rapidly finding a good approximate solution, while allowing users to weight the criteria as needed. The heuristic is evaluated with Pufferscale, a new, generic rescaling manager for microservice-based distributed storage systems [18].

To validate our approach in a real-world ecosystem, we showcase the use of Pufferscale as a means to enable storage malleability in the HEPnOS storage system for high energy physics applications.

7.2. Cloud and Edge processing

7.2.1. Benchmarking Edge processing frameworks

Participants: Pedro de Souza Bento Da Silva, Alexandru Costan, Gabriel Antoniu.

With the spectacular growth of the Internet of Things, edge processing emerged as a relevant means to offload data processing and analytics from centralized Clouds to the devices that serve as data sources (often provided with some processing capabilities). While a large plethora of frameworks for edge processing were recently proposed, the distributed systems community has no clear means today to discriminate between them. Some preliminary surveys exist, focusing on a feature-based comparison.

We claim that a step further is needed, to enable a performance-based comparison. To this purpose, the definition of a benchmark is a necessity. We make a step towards the definition of a methodology for benchmarking Edge processing frameworks [20].

7.2.2. Analytical models for performance evaluation of stream processing

Participants: José Aguilar Canepa, Pedro de Souza Bento Da Silva, Alexandru Costan, Gabriel Antoniu.

One of the challenges of enabling the Edge computing paradigm is to identify the situations and scenarios in which Edge processing is suitable to be applied. To this end, applications can be modeled as a graph consisting of tasks as nodes and data dependencies between them as edges. The problem comes down to deploying the application graph onto the network graph, that is, operators need to be put on machines, and finding the optimal cut in the graph between the Edge and Cloud resources (i.e., nodes in the network graph).

We have designed an algorithm that finds the optimal execution plan, with a rich cost model that lets users to optimize whichever goal they might be interested in, such as monetary costs, energetic consumption or network traffic, to name a few.

In order to validate the cost model and the effectiveness of the algorithm, a series of experiments were designed using two real-life stream processing applications: a closed-circuit television surveillance system, and an earthquake early warning system.

Two network infrastructures were designed to run the applications. The first one is a state-of-art infrastructure where all processing is done on the Cloud to serve as benchmark. The second one is an infrastructure produced by the algorithm. Both scenarios were executed on the Grid'5000. Several experiments are currently underway. The trade-offs of executing Cloud/Edge workloads with this model were published in [19].

7.2.3. Modeling smart cities applications

Participants: Edgar Romo Montiel, Pedro de Souza Bento Da Silva, Alexandru Costan, Gabriel Antoniu.

Smart City applications have particular characteristics in terms of data processing and storage, which need to be taken into account by the underlying serving layers. The objective of this new activity is to devise clear models of the data handled by such applications. The data characteristics and the processing requirements does not have to match one-to-one. In some cases, some particular types of data might need one or more types of processing, depending on the use case. For example, small and fast data coming from sensors do not always have to be processed in real-time, but they could also be processed in a batch manner at a later stage.

This activity is the namely the topic of the SmartFastData associated team with the Instituto Politécnico Nacional of Mexico.

In a first phase, we focused on modeling the stream rates of data from sets of sensors in Smart Cities, specifically, from vehicles inside a closed coverage area. Those vehicles are connected in a V2I VANET, and they interact to applications in the Cloud such as traffic reports, navigation apps, multimedia downloading etc. This led to the design of a mathematical model to predict the time that a mobile sensor resides within a geographical designated area.

The proposed model uses Coxian distributions to estimate the time a vehicle requests Cloud services, so that the core challenge is to adjust their parameters. It was achieved by validating the model against real-life data traces from the City of Luxembourg, through extensive experiments on the Grid'5000.

Next, these models were used to estimate the resources needed in the Cloud (or at the Edge) in order to process the whole stream of data. We designed an auto-Scaling module able to adapt the resources with respect to the load. Using the Grid'5000, we evaluated the various possibility to place the prediction module: *(i)* at the Edge, close to data with less accuracy but faster results; or *(ii)* in the Cloud, with higher accuracy due to the global data, but higher latency as well.

7.3. AI across the digital continuum

7.3.1. Machine Learning in the context of Edge stream processing.

Participants: Pedro de Souza Bento Da Silva, Alexandru Costan, Gabriel Antoniu.

Our research aims to improve the accuracy of Earthquake Early Warning (EEW) systems by means of machine learning. EEW systems are designed to detect and characterize medium and large earthquakes before their damaging effects reach a certain location.

Traditional EEW methods based on seismometers fail to accurately identify large earthquakes due to their sensitivity to the ground motion velocity. The recently introduced high-precision GPS stations, on the other hand, are ineffective to identify medium earthquakes due to its propensity to produce noisy data. In addition, GPS stations and seismometers may be deployed in large numbers across different locations and may produce a significant volume of data consequently, affecting the response time and the robustness of EEW systems.

In practice, EEW can be seen as a typical classification problem in the machine learning field: multi-sensor data are given in input, and earthquake severity is the classification result. We introduce the Distributed Multi-Sensor Earthquake Early Warning (DMSEEW) system, a novel machine learning-based approach that combines data from both types of sensors (GPS stations and seismometers) to detect medium and large earthquakes.

DMSEEW is based on a new stacking ensemble method which has been evaluated on a real-world dataset validated with geoscientists. The system builds on a geographically distributed infrastructure (deployable on clouds and edge systems), ensuring an efficient computation in terms of response time and robustness to partial infrastructure failures. Our experiments show that DMSEEW is more accurate than the traditional seismometer-only approach and the combined-sensors (GPS and seismometers) approach that adopts the rule of relative strength.

These results have been accepted for publication at AAAI, a "A*" conference in the area of Artificial Intelligence [21].

7.3.2. ZettaFlow: Unified Fast Data Storage and Analytics Platform for IoT

Participants: Ovidiu-Cristian Marcu, Alexandru Costan, Gabriel Antoniu.

The ZettaFlow platform (system of systems) provides a high-performance multi-model analytics-oriented storage and processing system, while supporting publish-subscribe streams and streaming, key-value and in-memory columnar APIs [16].

The ZettaFlow project is funded by EIT Digital from October 2019 to December 2020. It includes three partners: Inria for the platform development, TU Berlin for edge to cloud IoT optimizations with microservices, and Systematic Paris Region for the go-to-market strategy.

Our goal is to create a startup that will commercialize the ZettaFlow platform: a dynamic, unified and auto-balanced real-time storage and analytics industrial IoT platform. ZettaFlow will provide real-time visibility into machines, assets and factory operations and will automate data driven decisions for high-performance industrial processes.

ZettaFlow will bring a threefold impact to the IoT market.

1. Enable novel real-time edge applications that truly automate manufacturing, transportation and utilities processes.
2. Reduce deployment efforts and time-to-decision of IoT edge-cloud applications by 75% through automation, unified dynamic data management and streaming analytics.
3. Reduce human costs for monitoring and engineering (through edge intelligence) and IoT hardware costs by 50% through unified data collection/storage/analytics.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR

8.1.1.1. OverFlow (2015–2019)

Participants: Alexandru Costan, Pedro de Souza Bento Da Silva, Paul Le Noac'h.

Project Acronym: OverFlow

Project Title: Workflow Data Management as a Service for Multisite Applications

Coordinator: Alexandru Costan

Duration: October 2015–October 2019

Other Partners: None (Young Researcher Project, JCJC)

External collaborators: Kate Keahey (University of Chicago and Argonne National Laboratory), Bogdan Nicolae (Argonne National Lab)

Web site: https://sites.google.com/view/anroverflow

This project investigates approaches to data management enabling an efficient execution of geographically distributed workflows running on multi-site clouds.

In 2019, we focused on the challenges of stream processing at the Edge. In particular, Edge computing presents a significant opportunity to realize the potential of distributed ML models with regards to low latency, high availability and privacy. It allows for instance inferences on simple image, video or audio classification; as only the final result is transmitted, delays are minimized, while privacy and bandwidth are preserved in IoT applications. Also, neural networks could be partitioned such that some layers are evaluated at the Edge and the rest in the cloud.

In this context we proposed an architecture in which the initial layers can be used for feature-abstraction functions: as data travels through the neural network, they abstract into high-level features, which are more lightweight, helping reduce latency.

8.1.2. Other National Projects

8.1.2.1. HPC-Big Data Inria Project Lab (IPL)

Participants: Gabriel Antoniu, Alexandru Costan, Daniel Rosendo, Pedro de Souza Bento Da Silva.

Project Acronym: HPC-BigData

Project Title: The HPC-BigData Inria Project Lab

Coordinator: Bruno Raffin

Duration: 2018–2022

Web site: https://project.inria.fr/hpcbigdata/

The goal of this HPC-BigData IPL is to gather teams from the HPC, Big Data and Machine Learning (ML) areas to work at the intersection between these domains. Research is organized along three main axes: high performance analytics for scientific computing applications, high performance analytics for big data applications, infrastructure and resource management. Gabriel Antoniu is a member of the Advisory Board and leader of the Frameworks work package.

In 2019, Daniel Rosendo, who was hired in the context of this IPL project, focused on assessing the state of the art in high performance analytics on hybrid HPC/Big Data infrastructure. In particular, a new path for future work was identified: running Machine Learning algorithm at the Edge.

8.1.2.2. ADT Damaris 2

Participants: Ovidiu-Cristian Marcu, Gabriel Antoniu, Luc Bougé.

Project Acronym: ADT Damaris

Project Title: Technology development action for the Damaris environment

Coordinator: Gabriel Antoniu

Duration: 2019–2021

Web site: https://project.inria.fr/damaris/

This action aims to support the development of the Damaris software. Inria's *Technological Development Office* (D2T, *Direction du Développement Technologique*) provided 2 years of funding support for a senior engineer.

Ovidiu Marcu has been funded through this project to document, test and extend the Damaris software and make it a safely distributable product. In 2019, the main goal was to add Big Data analytics support in Damaris. We have extended Damaris with a streaming interface for writing and analyzing in real-time simulation data through KerA, a distributed streaming storage system.

KerA is further coupled with RAMCloud for in-memory key-value transactions and with Apache Flink for streaming analytics in an architecture that leverages Apache Arrow as in-memory columnar data representation for co-located streaming. This hybrid HPC-Big Data architecture is subject to further exploration within the ZettaFlow.io startup.

8.1.2.3. 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. Collaborations in European Programs, Except FP7 & H2020

8.2.1.1. ZettaFlow: Unified Fast Data Storage and Analytics Platform for IoT

Program: EIT Digital Innovation Factory

Project acronym: ZettaFlow

Project title: ZettaFlow: Unified Fast Data Storage and Analytics Platform for IoT

Duration: October 2019–December 2020

Technical Coordinator: Ovidiu Marcu

Other partners: Technische Universität Berlin and System@tic

Web site: https://zettaflow.io/

The objective of this project is to create a startup in order to commercialize the ZettaFlow platform: a dynamic, unified and auto-balanced real-time storage and analytics industrial IoT platform. ZettaFlow is based on KerA, a streaming storage system prototype developed within the KerData team. ZettaFlow will provide real-time visibility into machines, assets and factory operations and will automate data driven decisions for high-performance industrial processes.

8.2.1.2. FlexStream: Automatic Elasticity for Stream-based Applications

Program: PHC PROCOPE 2020
Project acronym: FlexStream
Project title: Automatic Elasticity for Stream-based Applications
Duration: January 2020–December 2021
Coordinator: Alexandru Costan
Other partners: University of Dusseldorf (UDUS)

Elasticity is one of the key features of cloud computing providing virtual resources as needed according to dynamically changing workloads. This allows to minimize costs and reduce time-to-decision of IoT edge-cloud applications. However, while the underlying resources may easily be scaled many applications and services are not designed to support elastic scalability or require an administrator to manually control elastic scaling.

This project aims at developing concepts providing automatic scaling for stream processing applications. In particular, FlexStream aims at developing and evaluating a prototype which will integrate a stream ingestion-system from IRISA and an in-memory storage from UDUS. For this approach a tight cooperation is mandatory in order to be successful which in turn requires visits on both sides and longer exchanges, especially for the involved PhD students, in order to allow an efficient integrated software design, development as well as joint experiments on large platforms and preparing joint publications.

8.2.2. Collaborations with Major European Organizations

8.2.2.1. BDVA and ETP4HPC

Gabriel Antoniu (as a working group leader) and Alexandru Costan (as a working group member) contributed to the new Strategic Research Agenda (version 4) of European Technology Platform in the area of High-Performance Computing (ETP4HPC).

Gabriel Antoniu and Alexandru Costan are serving as Inria representatives in the working group dedicated to *HPC-Big Data* convergence within the Big Data Value Association (BDVA).

8.2.2.2. International Initiatives

8.2.2.2.1. 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 series of invitation-based international workshops.

In 2019 Gabriel Antoniu was invited again to contribute to the second and third workshops of the BDEC2 series, where he presented two white papers on HPC-Big Data convergence at the level of data processing.

8.3. International Initiatives

8.3.1. Inria International Labs

8.3.1.1. UNIFY: An associated team involved in the JLESC international lab

Title: UNIFY: Intelligent Unified Data Services for Hybrid Workflows Combining Compute-Intensive Simulations and Data-Intensive Analytics at Extreme Scales
Inria International Lab: JLESC: Joint Laboratory for Extreme Scale Computing
International Partner: Argonne National Laboratory (USA) — Department of Mathematics, Symbolic Computation Group — Tom Peterka
Start year: 2019
See also: https://team.inria.fr/unify

The landscape of scientific computing is being radically reshaped by the explosive growth in the number and power of digital data generators, ranging from major scientific instruments to the Internet of Things (IoT) and the unprecedented volume and diversity of the data they generate. This requires a rich, extended ecosystem including simulation, data analytics, and learning applications, each with distinct data management and analysis needs.

Science activities are beginning to combine these techniques in new, large-scale workflows, in which scientific data is produced, consumed, and analyzed across multiple distinct steps that span computing resources, software frameworks, and time. This paradigm introduces new data-related challenges at several levels.

The UNIFY Associate Team aims to address three such challenges. First, to allow scientists to obtain fast, real-time insight from complex workflows combining extreme-scale computations with data analytics, we will explore how recently emerged Big Data processing techniques (e.g., based on stream processing) can be leveraged with modern in situ/in transit processing approaches used in HPC environments.

Second, we will investigate how to use transient storage systems to enable efficient, dynamic data management for hybrid workflows combining simulations and analytics.

Finally, the explosion of learning and AI provides new tools that can enable much more adaptable resource management and data services than available today, which can further optimize such data processing workflows.

8.3.2. Inria Associate Teams Not Involved in an Inria International Labs

8.3.2.1. SmartFastData

Title: Efficient Data Management in Support of Hybrid Edge/Cloud Analytics for Smart Cities

International Partner: Instituto Politécnico Nacional (Mexico) — Centro de Investigación en Computación — Rolando Menchaca-Mendez

- Start year: 2019
- See also: https://team.inria.fr/smartfastdata/

The proliferation of small sensors and devices that are capable of generating valuable information in the context of the Internet of Things (IoT) has exacerbated the amount of data flowing from all connected objects to private and public cloud infrastructures. In particular, this is true for Smart City applications, which cover a large spectrum of needs in public safety, water and energy management. Unfortunately, the lack of a scalable data management subsystem is becoming an important bottleneck for such applications, as it increases the gap between their I/O requirements and the storage performance.

The vision underlying the SmartFastData associated team is that, by smartly and efficiently combining the data-driven analytics at the edge and in the cloud, it becomes possible to make a substantial step beyond state-of-the-art prescriptive analytics through a new, high-potential, faster approach to react to the sensed data.

The goal is to build a data management platform that will enable comprehensive joint analytics of past (historical) and present (real-time) data, in the cloud and at the edge, respectively, allowing to quickly detect and react to special conditions and to predict how the targeted system would behave in critical situations.

In 2019, the first objective of the associated team (i.e., exploring analytical models for performance evaluation of stream storage and ingestion systems) was achieved by means of the two internships of José Canepa and Edgar Romo (described in the New Results section) as well as the visit of Mario Rivero as an Invited Professor, who set up the main research agenda for those internships.

8.4. International Research Visitors

8.4.1. Visits of International Scientists

Rosa Badia: Barcelona Supercomputing Center, Spain. Dates: 13-14 March 2019

Michael Schottner: University of Dusseldorf, Germany. Dates: 13-15 March 2019

Valentin Cristea: Politehnica University of Bucharest, Romania. Dates: 13-15 March 2019

Toni Cortés: Universitat Politècnica Catalunya, Spain. Dates: 4-5 November 2019

Kate Keahey: Argonne National Lab, USA. Dates: 4-5 November 2019

Matthieu Dorier: Argonne National Lab, USA. Dates: 4-5 November 2019

8.4.1.1. Invited Professors

Mario Rivero (Professor, Instituto Politécnico Nacional, Mexico) was an invited professor in the KerData team from June to July 2019, through the *Scientist Invitation Program* of IRISA and ISTIC. During his stay, he gave several talks at Inria/IRISA and worked on the modeling Smart City applications, laying the path for the work program of the upcoming internships of José Aguilar-Canepa and Edgar Romo.

8.4.1.2. Internships

Jose Aguilar-Canepa (PhD student, Instituto Politécnico Nacional, Mexico) has done a 3-month internship within the team, working with Alexandru Costan and Pedro Silva on hybrid Edge/Cloud stream processing. This work is validated through large scale experiments on Grid'5000 and is subject to a journal paper in submission, currently on the works, to be submitted by January 2020.

Edgar Romo (PhD student, Instituto Politécnico Nacional, Mexico) did a 3-month internship at KerData from September to November 2019. He worked on Objective 2 of the SmartFastData Associate Team, specifically on designing a complex model for predicting the stream arrival rates for vehicular networks in Smart Cities. To validate this proposal, he carried out several experiments on Grid'5000; this work is currently the topic of a workshop paper submission.

8.4.2. Visits to International Teams

Alexandru Costan and Gabriel Antoniu visited the NDS-Lab team at Instituto Politécnico Nacional from October 24 to November 3, 2019, in the context of the SmartFastData associate team. Working closely with Rolando Menchaca, they defined the work program for the upcoming year with respect to the team's objectives. They also presented KerData's vision on future hybrid analytics combining Edge, Cloud and HPC computing.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Organization

9.1.1.1. General Chair, Scientific Chair

Luc Bougé: Steering Committee Chair of the Euro-Par International Conference on Parallel and Distributed Computing. Euro-Par celebrated its 25th anniversary in Göttingen, Germany, this year.

Gabriel Antoniu: Co-Chair of Conv'2019 - the HPC-AI-Big Data Convergence Days.

9.1.1.2. Member of the Organizing Committees

Gabriel Antoniu: Member of the Organizing Committee of Conv'2019, the HPC-AI-Big Data Convergence Days.

9.1.2. Scientific Events: Selection

9.1.2.1. Chair of Conference Program Committees

Gabriel Antoniu: Track Co-Chair for the Clouds and Distributed Systems Track of the ACM/IEEE SC19 conference, Denver, USA.

Alexandru Costan: Program Co-Chair of the ScienceCloud 2019 international workshop held in conjunction with ACM HPDC 2019, Phoenix, AZ, USA.

9.1.2.2. Member of the Conference Program Committees

Alexandru Costan: IEEE/ACM SC'19 (Technical Program: Cloud Track, Posters and ACM Student Research Competition), ACM/IEEE CCGrid 2019, IEEE Cluster 2019, IEEE/ACM UCC 2019, IEEE Big Data 2019, SCRAMBL 2019, CSCS 2019, CEBDA 2019, CCIW 2019, IEEE CSE 2018, IEEE CloudCom 2019.

Gabriel Antoniu: IEEE Cluster 2019, IEEE IPDPS 2019, STREAM-ML 2019.

9.1.2.3. Reviewer

Alexandru Costan: ACM HPDC 2019, IEEE IPDPS 2019.

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

Gabriel Antoniu: Associate Editor of JPDC, the Elsevier Journal of Parallel and Distributed Computing.

9.1.3.2. Reviewer - Reviewing Activities

Alexandru Costan: IEEE Transactions on Parallel and Distributed Systems, Future Generation Computer Systems, Concurrency and Computation Practice and Experience, IEEE Transactions on Cloud Computing, Journal of Parallel and Distributed Computing.

Gabriel Antoniu: SoftwareX, Philosophical Transactions A.

9.1.4. Invited Talks

Gabriel Antoniu

- Fourth BDEC2 Workshop on Big Data and Extreme Computing, San Diego, in October 2019. *Towards a demonstrator of the Sigma Data Processing Architecture for BDEC 2*. URL: https://www.exascale.org/bdec/agenda/poznan.
- Third BDEC2 Workshop on Big Data and Extreme Computing, Poznan, in May 2019. *ZettaFlow: Towards High-Performance ML-based Analytics across the Digital Continuum*. URL: https://www.exascale.org/bdec/agenda/sandiego.
- Ninth Workshop of the Joint Laboratory for Extreme-Scale Computing (JLESC) in April 2019. *Scalable Data Ingestion for Stream Processing*. URL: https://jlesc.github.io/.

Alexandru Costan

- Inria/ IPN Joint Workshop, Mexico City, November 2019. *From Big Data to Fast Data: Efficient Stream Data Managements*.

9.1.5. Leadership within the Scientific Community

Luc Bougé: Co-Vice-President of the *French Society for Informatics* (*Société informatique de France*, SIF), in charge of the Teaching Department.

Gabriel Antoniu

ETP4HPC Since 2019, co-leader of the working group on Programming Environments and co-lead of two research clusters, contributing to the next Strategic Research Agenda of ETP4HPC (to appear).

International lab management *Vice Executive Director of JLESC* for Inria. JLESC is the Joint Inria-Illinois-ANL-BSC-JSC-RIKEN/AICS Laboratory for Extreme-Scale Computing. Within JLESC, he also serves as a *Topic Leader* for Data storage, I/O and in situ processing for Inria.

Team management *Head of the KerData Project-Team* (Inria-ENS Rennes-INSA Rennes).

International Associate Team management Leader of the UNIFY Associate Team with Argonne National Lab (2019–2021).

Technology development project management Coordinator of the Damaris ADT project (2016–2018), to be continued with the Damaris 2 ADT project (2019–2021).

9.1.6. Scientific Expertise

Luc Bougé

HCERES: Chair of the evaluation committee for the CRIStAL joint laboratory, Lille (report)

HCERES: Member of the evaluation committee for the IRCICA research institute, Lille

9.1.7. Research Administration

Luc Bougé

ANR: Member of the management team for the IA for Humanity governmental program launched in March 2018. Bertrand Braunschweig, Inria, is the scientific director of the program.

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: Scalable Distributed Systems, 10 hours (lectures), M1 level, SDS Module, EIT ICT Labs Master School, France.
- Master: Infrastructures for Big Data, 10 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.

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, 28 hours (lectures, lab sessions), M2, INSA Rennes.
- Master: Big Data Project, 28 hours (project), M2, INSA Rennes.

Luc Bougé

- Bachelor: Introduction to programming concepts, 36 hours (lectures), L3 level, Informatics program, ENS Rennes, France.
- Bachelor: Introduction to scientific research, 24 hours. Research center visits, individual research project supervised by local researchers, student seminars, summer internships, etc.
- Master Program, Rennes: Invited presentation to the M2 students about *Preparing your applications after your PhD* (October 2019); *Informatics as a scientific activity: Towards a responsible research* (November 2019).

Pedro Silva

- Master: Algorithms for Big Data, 4 hours (lectures), M2, INSA Rennes.
- Master: Algorithms for Big Data, 6 hours (lab sessions), M2, INSA Rennes.

9.2.2. Supervision

9.2.2.1. HdR completed this year

Alexandru Costan: *From Big Data to Fast Data: Efficient Stream Data Management*, ENS Rennes, March 2019 [15].

9.2.2.2. PhD completed this year

Nathanaël Cheriere: *Towards Malleable Distributed Storage Systems? From Models to Practice*, ENS Rennes, thesis defended in November 2019, co-advised by Gabriel Antoniu and Matthieu Dorier [14].

9.2.2.3. PhD in progress

Daniel Rosendo: *Enabling HPC-Big Data Convergence for Intelligent Extreme-Scale Analytics*, INSA Rennes, thesis started in October 2019, co-advised by Gabriel Antoniu, Alexandru Costan and Patrick Valduriez (Inria).

Paul Le Noac'h: *Workflow Data Management as a Service for Multi-Site Applications*, INSA Rennes, thesis started in November 2016, co-advised by Luc Bougé and Alexandru Costan. Thesis stopped in February 2019.

9.2.3. Juries

Luc Bougé: Member of the jury the *CAPES of mathématiques*, *Informatics track*. This national committee selects more than 1000 mathematics teachers per year for French secondary schools and high-schools.

9.3. Popularization

9.3.1. Internal or external Inria responsibilities

Alexandru Costan

- In charge of internships at the Computer Science Department of INSA Rennes.
- In charge of the organization of the IRISA D1 Department Seminars.

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'IO: 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^{o} 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] N. CHERIERE.*Towards Malleable Distributed Storage Systems: From Models to Practice*, École normale supérieure de Rennes, November 2019, https://tel.archives-ouvertes.fr/tel-02376032

[15] A. COSTAN.*From Big Data to Fast Data: Efficient Stream Data Management*, ENS Rennes, March 2019, Habilitation à diriger des recherches, https://hal.archives-ouvertes.fr/tel-02059437

Invited Conferences

[16] G. ANTONIU, A. COSTAN, O.-C. MARCU.*ZettaFlow: Towards High-Performance ML-based Analytics across the Digital Continuum*, in "BDEC2 2019 - Workshop on Big Data and Extreme-scale Computing",

San Diego, United States, San Diego Supercomputing Center, October 2019, 4, https://hal.archives-ouvertes.fr/hal-02428382

[17] G. ANTONIU, A. COSTAN, O.-C. MARCU, M. HERNÁNDEZ-PÉREZ, N. STOJANOVIC.*Towards a demonstrator of the Sigma Data Processing Architecture for BDEC 2*, in "BDEC2 2019 - Workshop on Big Data and Extreme-scale Computing", Poznan, Poland, Poznan Supercomputing and Networking Center, May 2019, 4, https://hal.archives-ouvertes.fr/hal-02428391

International Conferences with Proceedings

[18] N. CHERIERE, M. DORIER, G. ANTONIU.*Is it Worth Relaxing Fault Tolerance to Speed Up Decommission in Distributed Storage Systems?*, in "CCGrid 2019 - IEEE/ACM International Symposium in Cluster, Cloud, and Grid Computing", Larnaca, Cyprus, IEEE, May 2019, p. 1-10 [*DOI :* 10.1109/CCGRID.2019.00024], https://hal.archives-ouvertes.fr/hal-02116727

[19] P. SILVA, A. COSTAN, G. ANTONIU.*Investigating Edge vs. Cloud Computing Trade-offs for Stream Processing*, in "BigData 2019 - IEEE International Conference on Big Data", Los Angeles, United States, IEEE, December 2019, https://hal.archives-ouvertes.fr/hal-02415684

[20] P. SILVA, A. COSTAN, G. ANTONIU.*Towards a Methodology for Benchmarking Edge Processing Frameworks*, in "IPDPSW 2019 - IEEE International Parallel and Distributed Processing Symposium Workshops", Rio de Janeiro, Brazil, IEEE, May 2019, p. 904-907 [*DOI :* 10.1109/IPDPSW.2019.00149], https://hal.inria.fr/hal-02310154

Other Publications

[21] K. FAUVEL, D. BALOUEK-THOMERT, D. MELGAR, P. SILVA, A. SIMONET, G. ANTONIU, A. COSTAN, V. MASSON, M. PARASHAR, I. RODERO, A. TERMIER.*A Distributed Multi-Sensor Machine Learning Approach to Earthquake Early Warning*, November 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02373429

References in notes

[22] *Amazon Elastic Map-Reduce (EMR)*, 2017, https://aws.amazon.com/emr/

[23] *Digital Single Market*, 2015, https://ec.europa.eu/digital-single-market/en/digital-single-market/

[24] *European Exascale Software Initiative*, 2013, http://www.eesi-project.eu/

[25] *The European Technology Platform for High-Performance Computing*, 2012, http://www.etp4hpc.eu/

[26] *European Cloud Strategy*, 2012, https://ec.europa.eu/digital-single-market/en/european-cloud-computing-strategy/

[27] *Apache Flink*, 2016, http://flink.apache.org/

[28] *International Exascale Software Program*, 2011, http://www.exascale.org/iesp/

[29] *Scientific challenges of the Inria Rennes-Bretagne Atlantique research centre*, 2016, https://www.inria.fr/centre-inria-rennes-bretagne-atlantique/

[30] *Inria's strategic plan "Towards Inria 2020"*, 2016, https://www.inria.fr/recherche-innovation/

[31] *Joint Laboratory for Extreme Scale Computing (JLESC)*, 2017, https://jlesc.github.io/

[32] *Apache Spark*, 2017, http://spark.apache.org/

[33] *Storm*, 2014, http://storm.apache.org/

[34] 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^{o} 11, p. 1033–1044

[35] J. DEAN, S. GHEMAWAT.*MapReduce: simplified data processing on large clusters*, in "Communications of the ACM", 2008, vol. 51, n^{o} 1, p. 107–113

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

Project-Team LACODAM

Creation of the Team: 2016 January 01, updated into Project-Team: 2017 November 01

Keywords:

Computer Science and Digital Science:

A2.1.5. - Constraint programming
A3.1.1. - Modeling, representation
A3.1.6. - Query optimization
A3.1.11. - Structured data
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
A3.4.1. - Supervised learning
A3.4.2. - Unsupervised learning
A3.4.6. - Neural networks
A3.4.8. - Deep learning
A9.1. - Knowledge
A9.2. - Machine learning
A9.3. - Signal analysis
A9.6. - Decision support
A9.7. - AI algorithmics
A9.8. - Reasoning

Other Research Topics and Application Domains:

B1.2.2. - Cognitive science
B2.3. - Epidemiology
B2.4.1. - Pharmaco kinetics and dynamics
B3.5. - Agronomy
B3.6. - Ecology
B3.6.1. - Biodiversity

1. Team, Visitors, External Collaborators

Research Scientists

Louis Bonneau de Beaufort [École nationale supérieure agronomique de Rennes, Researcher]
Luis Galárraga Del Prado [Inria, Researcher]

Faculty Members

Alexandre Termier [Team leader, Univ de Rennes I, Professor, HDR]

Tassadit Bouadi [Univ de Rennes I, Associate Professor, from Jul 2019]
Elisa Fromont [Univ de Rennes I, Professor, HDR]
Thomas Guyet [École nationale supérieure agronomique de Rennes, Associate Professor]
Christine Largouët [École nationale supérieure agronomique de Rennes, Associate Professor, HDR]
Véronique Masson [Univ de Rennes I, Associate Professor]
Laurence Rozé [INSA Rennes, Associate Professor]

External Collaborators
Johanne Bakalara [Univ de Rennes I]
Philippe Besnard [CNRS, HDR]
Romaric Gaudel [École nationale de la statistique et de l'analyse de l'information]
Raphael Gauthier [INRA]
Anne-Isabelle Graux [INRA]

Technical Staff
Remi Adon [Inria, Engineer, from Dec 2019]

PhD Students
Erwan Bourrand [Advisor SLA, PhD Student]
Kevin Fauvel [Inria, PhD Student]
Samuel Felton [Univ de Rennes I, PhD Student, from Oct 2019]
Camille Sovanneary Gauthier [Louis Vuitton, PhD Student, from Apr 2019]
Maël Guillemé [Eniergency, PhD Student, granted by CIFRE]
Colin Leverger [Orange Labs, PhD Student]
Gregory Martin [PSA Automobiles, PhD Student, granted by CIFRE]
Anh Duong Nguyen [Vietnam, PhD Student, until May 2019]
Alban Siffer [AMOSSYS, PhD Student, granted by CIFRE]
Antonin Voyez [Inria, PhD Student, from Dec 2019]
Yichang Wang [CSC Grant, PhD Student]
Heng Zhang [Atermes, PhD Student, granted by CIFRE]

Visiting Scientist
Alexandre Sahuguede [Univ Paul Sabatier, until Jan 2019]

Administrative Assistant
Gaelle Tworkowski [Inria, Administrative Assistant]

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 stream 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.

The objective of the LACODAM 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. Our approaches are mostly dedicated to provide novel tools to data scientists, that can either performs tasks not addressed by any other tools, or that improve the performance in some area for existing tasks (for instance reducing execution time, improving accuracy or better handling imbalanced data).

3. Research Program

3.1. Introduction

The three original research axes of the LACODAM project-team are the following. First, we briefly introduce these axes, as well as their interplay. We then introduce the axis of *Interpretable AI* (Section 3.4), whose emergence is a response to the current societal needs.

- The first research axis (Section 3.2) 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 overall goals.
- The second axis (Section 3.3) 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 (Section 3.4) 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.

Figure 1 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.

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 data mining approaches are based on agnostic approaches designed to cope with pattern explosion, we propose to develop data mining techniques that rely on knowledge-based artificial intelligence techniques. This entails the use of structured knowledge representations, as well as reasoning methods, in combination with mining.

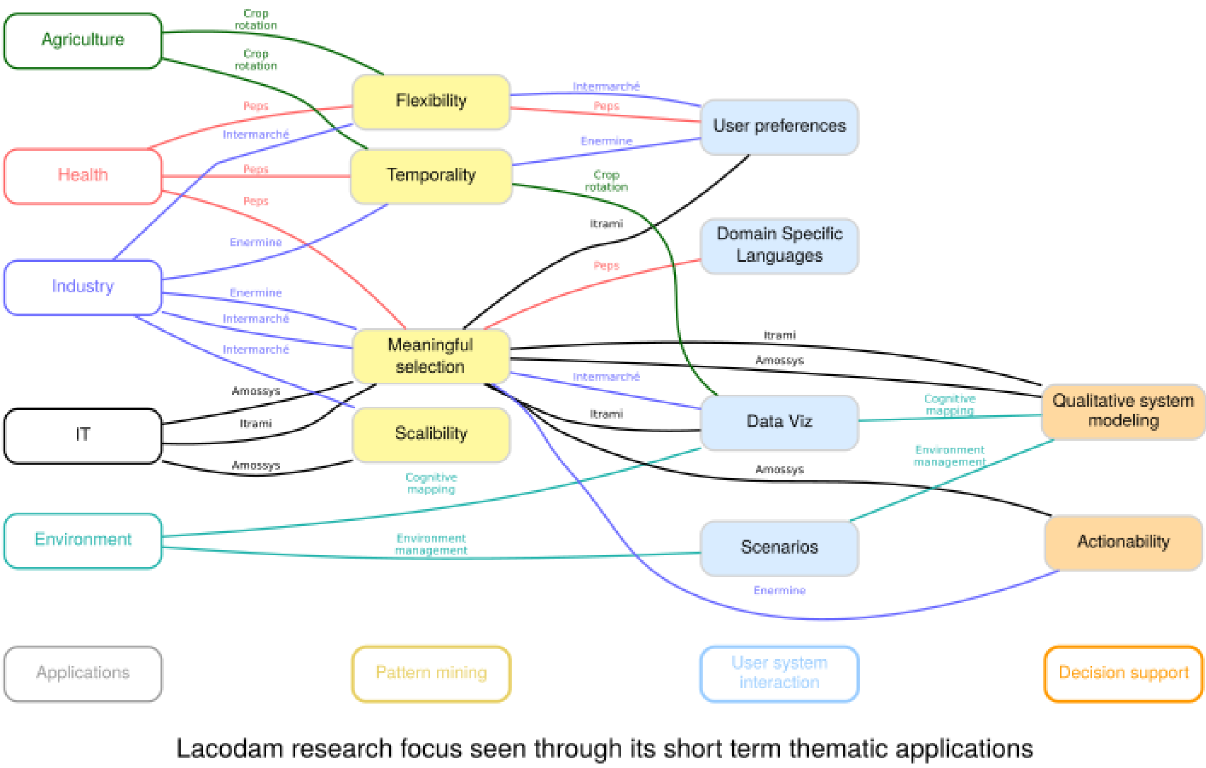


Figure 1. LACODAM research topics organized by axis and application

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 to 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

Patterns have proved to be quite useful for decision-aid. Predictive sequential patterns, to give an example, have a direct application in diagnosis. Itemsets and contrast patterns can be used for interpretable machine learning (ML). In regards to diagnosis, LACODAM inherits, from the former DREAM team, a strong background in decision support systems with internationally recognized expertise in this field. This subfield of AI (Artificial Intelligence) 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 (e.g., a ML classifier), a living being, etc. In relation to interpretable machine learning (ML), this subfield is concerned with the conception of models whose answers are understandable by users. This can be achieved by inducing inherently white-box models from data such as rule-based classifiers/regressors, or by mining rules and explanations from black-box models. The latter setting is quite common due to the high accuracy of black-box models compared to natively interpretable models. Pattern mining is a powerful tool to mine explanations from black-box systems. Those explanations can be used to diagnose biases in systems, either to debug and improve the model, or to generate trust in the verdicts of intelligent software agents.

The increasing volumes of data coming from a range of different systems (ex: sensor data from agro-environmental systems, log data from software systems and ML models, 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.

Three 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.
- *Mining explanations from ML systems.* Interpretable ML and AI is a current trend for technical, ethical, and legal reasons [27]. In this regard, pattern mining can be used to spot regularities that arise when a complex black-box model yields a particular verdict. For instance, one may want to know the conditions under which the control module of a self-driving car decided to stop without apparent reason, or which factors caused a ML-based credit assessor to reject a loan request. Patterns and conditions are the building blocks for the generation of human-readable explanations for such black-box systems.

3.5. Interpretability

The pervasiveness of complex decision support systems, as well as the general consensus about the societal importance of understanding the rationale embedded in such systems [0], has given momentum to the field of interpretable ML. Being a team specialized in data science, we are fully aware that many problems can be solved by means of complex and accurate ML models. Alas, this accuracy sometimes comes at the expense of interpretability, which can be a major requirement in some contexts (e.g., regression using expertise/rule mining). For this reason, one of the interests of LACODAM is the study of the interpretability-accuracy trade-off. Our studies may be able to answer questions such as “how much accuracy can a model lose (or perhaps gain) by becoming more interpretable?”. Such a goal requires us to define interpretability in a more principled

[0] General Data Protection Regulation, recital 71 http://www.privacy-regulation.eu/en/r71.htm

way —an endeavour that has been very recently addressed, still not solved. LACODAM is interested in the two main currents of research in interpretability, namely the development of natively interpretable methods, as well as the construction of interpretable mediators between users and black-box models, known as post-hoc interpretability.

We highlight the link between interpretability and LACODAM's axes of decision support, and user/system interaction. In particular, interpretability is a prerequisite for proper user/system interaction and is a central incentive for the advent of data visualization techniques for ML models. This convergence has motivated our interest in *user-oriented post-hoc interpretability*, a sub-field of interpretable ML that adds the user into the formula when generating proper explanations of black-box ML algorithms. This rationale is supported by existing work [28] that suggests that interpretability possesses a subjective component known as plausibility. Moreover, our user-oriented vision meets with the notion of semantic interpretability, where an explanation may resort to high level semantic elements (objects in image classification, or verbal phases in natural language processing) instead of surrogate still-machine-friendly features (such as super-pixels). LACODAM will tackle all these unaddressed aspects of interpretable ML with other Inria teams through the IPL HyAIAI.

3.6. Long-term goals

The following perspectives are at the convergence of the four 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. Energiency, Amossys).

- **Resource Consumption Analysis for Optimizing Energy Consumption and Practices in Industrial Factories (Energiency)**. In order to increase their understanding of factory operation, 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 CIFRE PhD thesis of Maël (defended 16/12/2019) aimed 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 collaborated on this topic with the Amossys company and the EMSEC team of Irisa through the co-supervision of the CIFRE PhD of Alban Siffer (located in the EMSEC team, defended 19/12/2019). Our goal was in particular to design novel anomaly detection methods making minimal assumptions on the data, and able to scale to real traffic volumes.
- **Car Sharing Data Analysis**. Peugeot-Citroën (PSA) group’s know-how encompasses all areas of the automotive industry, from production to distribution and services. Among others, its aim is to provide a car sharing service in many large cities. This service consists in providing a fleet of cars and a “free floating” system that allows users to use a vehicle, then drop it off at their convenience in the city. To optimize their fleet and the availability of the cars throughout the city, PSA needs to analyze the trajectory of the cars and understand the mobility needs and behavior of their users. We tackle this subject together through the CIFRE PhD of Gregory Martin.

- **Multimodal Data Analysis for the Supervision of Sensitive Sites.** ATERMES is an international mid-sized company with a strong expertise in high technology and system integration from the upstream design to the long-life maintenance cycle. It has recently developed a new product, called BARIER TM("Beacon Autonomous Reconnaissance Identification and Evaluation Response"), which provides operational and tactical solutions for mastering borders and areas. Once in place, the system allows for a continuous night and day surveillance mission with a small crew in the most unexpected rugged terrain. The CIFRE PhD of Heng Zhang aims at developing a deep learning architecture and algorithms able to detect anomalies (mainly persons) from multimodal data. The data are "multimodal" because information about the same phenomenon can be acquired from different types of detectors, at different conditions, in multiple experiments.
- **Root Cause Analysis in Networks**. AdvisorSLA is a French company specialized in software solutions for network monitoring. For this purpose, the company relies on techniques of network metrology. By continuously measuring the state of the network, monitoring solutions detect events (e.g., overloaded router) that may degrade the network's operation and the quality of the services running on top of it (e.g., video transmission could become choppy). When a monitoring solution detects a potentially problematic sequence of events, it triggers an alarm so that the network manager can take actions. Those actions can be preventive or corrective. Some statistics show that only 40% of the triggered alarms are conclusive, that is, they manage to signal a well-understood problem that requires an action from the network manager. This means that the remaining 60% are presumably false alarms. While false alarms do not hinder network operation, they do incur an important cost in terms of human resources. Thus, the CIFRE PhD of Erwan Bourrand proposes to characterize conclusive and false alarms. This will be achieved by designing automatic methods to "learn" the conditions that most likely precede each type of alarm, and therefore predict whether the alarm will be conclusive or not. This can help adjust existing monitoring solutions in order to improve their accuracy. Besides, it can help network managers automatically trace the causes of a problem in the network.

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 [31], [30] or medical devices [23], 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 algorithms 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 [0] 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

[0] PEPS: Pharmaco-Epidémiologie et Produits de Santé – Pharmacoepidemiology of health products

past years. The main characteristics of this data warehouse are described in [29]. 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.

- **Care Sequences for the Exploration of Medico-administrative Data**. The difficulty of analyzing medico-administrative data is the semantic gap between the raw data (for example, database record about the delivery at date t of drug with ATC 2 code N 02BE01) and the nature of the events sought by clinicians ("was the patient exposed to a daily dose of paracetamol higher than 3g?"). The solution that is used by epidemiologists consists in enriching the data with new types of events that, on the one side, could be generated from raw data and on the other side, have a medical interpretation. Such new abstract events are defined by clinician using proxies. For example, drugs deliveries can be translated in periods of drug exposure (drug exposure is a time-dependent variable for non-random reasons) or identify patient stages of illness, etc. A proxy can be seen as an abstract description of a care sequence.

  Currently, the clinicians are limited in the expression of these proxies both by the coarse expressivity of their tools and by the need to process efficiently large amount of data. From a semantic point of view, care sequences must fully integrate the temporal and taxonomic dimensions of the data to provide significant expression power. From a computational point of view, the methods employed must make it possible to efficiently handle large amounts of data (several millions care pathways). The aim of the PhD of Johanne Bakalara is to study temporal models of sequences in order 1) to show their abilities to specify complex proxies representing care sequences needed in pharmaco-epidemiological studies and 2) to build an efficient querying tool able to exploit large amount of care pathways.

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 such as 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 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.
- **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 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.

4.5. Education

- **Data-oriented Academic Counseling.** Course selection and recommendation are important aspects of any academic counseling system. The Learning Analytics community has long supported these activities via automatic, data-based tools for recommendation and prediction. LACODAM, in collaboration with the Ecuadorian research center CTI [0] has contributed to this body of research with the design of a tool that allows students to select multiple courses and predict their academic performance based on historical academic data. The tool resorts to interpretable machine learning techniques, and is intended to be used by the students before the counseling sessions to plan their upcoming semester at the Ecuadorian university ESPOL. The data visualization aspects of the tool, as well as the data science considerations and our emphasis on explainability are compiled in a paper that is under revision at the Learning Analytics and Knowledge Conference (LAK'20). The data science component of the tool was developed during the M1 internship of Mohammad Poul-Doust.

4.6. Others

- **RDF Archiving.** The dynamicity of the Semantic Web has motivated the development of solutions for RDF archiving, i.e., the task of storing and querying all previous versions of an RDF dataset. Notwithstanding the value of RDF archiving for data maintainers and consumers, this field of research remains under-developed for multiple reasons. These include notably (i) the lack of usability and scalability of the existing systems, (ii) no archiving support for multi-graph RDF datasets, (iii) the absence of a standard SPARQL extension for RDF archives, and (iv) a disregard of the evolution patterns of RDF datasets. The PhD thesis of Olivier Pelgrin aims at developing techniques towards a scalable and full-fledged archiving solution for the Semantic Web. This PhD thesis is a collaboration between LACODAM and the DAISY team at Aalborg University.

5. Highlights of the Year

5.1. Highlights of the Year

- Elisa Fromont was awarded a Junior Member position at the Institut Universitaire de France (IUF). This is a prestigious position given for 5 years (2019-2024), the selection process is especially competitive.
- Tassadit Bouadi (MCF Univ Rennes 1) joined the team in July 2019. Her research topics are skyline queries and preference mining. Her work will especially contribute to the design of approaches having results easier to grasp by human users.

6. New Software and Platforms

6.1. REMI

Mining Intuitive Referring Expressions in Knowledge Bases

[0] Centro de Tecnologías de Información, http://cti.espol.edu.ec/

KEYWORDS: RDF - Knowledge database - Referring expression

FUNCTIONAL DESCRIPTION: REMI takes an RDF knowledge base stored as an HDT file, and a set of target entities and returns a referring expression that is intuitive, i.e., the user is likely to understand it.

- Contact: Luis Galarraga Del Prado
- URL: http://gitlab.inria.fr/lgalarra/remi

6.2. HIPAR

Hierarchical Interpretable Pattern-aided Regression

KEYWORDS: Regression - Pattern extraction

FUNCTIONAL DESCRIPTION: Given a (tabular) dataset with categorical and numerical attributes, HIPAR is a Python library that can extract accurate hybrid rules that offer a trade-off between (a) interpretability, (b) accuracy, and (c) data coverage.

- Contact: Luis Galarraga Del Prado
- URL: https://gitlab.inria.fr/opelgrin/hipar

6.3. PyChronicle

KEYWORDS: Sequence - Sequential patterns - Pattern matching

FUNCTIONAL DESCRIPTION: Python library containing classes for representing sequences and chronicles, ie a representation of a temporal pattern. It implements efficient recognition algorithms to match chronicles in a long sequence.

- Participant: Thomas Guyet
- Contact: Thomas Guyet
- Publication: Énumération des occurrences d'une chronique
- URL: https://gitlab.inria.fr/tguyet/pychronicles

7. New Results

7.1. Introduction

In this section, we organize the bulk of our contributions this year along two of our research axes, namely Pattern Mining and Decision Support. Some other contributions lie within the domain of machine learning.

7.1.1. Pattern Mining

In the domain of pattern mining we can categorize our contributions along the following lines:

- *Efficient Pattern Mining (Sections 7.2-7.4).* In [9], we propose a method to accelerate itemset sampling on FPGAs, whereas [18] proposes SSDPS, an efficient algorithm to mine discriminant patterns in two-class datasets, common in genetic data. Finally [11] presents a succinct data structure that represents concisely a cube of skypatterns.
- *Semantics of Pattern Mining (Sections 7.5-7.6).*[14] discusses the ambiguity of the semantics of pattern mining with absent events (negated statements). Likewise [8] shows formal properties of admissible generalizations in pattern mining and machine learning.

7.1.2. Decision Support

In regards to the axis of decision support, our contributions can be organized in two categories: forecasting & prediction, and modelisation.

- *Forecasting & Prediction (Sections 7.7-7.9).* In [10], we propose solutions to automate the task of capacity planning in the context of a large data network as the one available at Orange. [17] applies machine learning techniques for estrus detection in diary farms. [21] proposes a machine learning architecture in multi-sensor environments for earthquake early warning.
- *Modelling (Section 7.10).* In [5] we present a modeling approach for the nutritional requirements of lactating sows.
- *Data Exploration (Section 7.11).*[6] proposes a formal framework for the exploration of care trajectories in medical databases.

7.1.3. Others

- *Machine Learning (Section 7.12-7.14).*[7], [16] proposes novel methods to optimize the F-measure in ML, and to improve the task of domain adaptation by source selection. [19] proposes the use of GANs to make time series classification more interpretable.

7.2. Accelerating Itemset Sampling using Satisfiability Constraints on FPGA

Finding recurrent patterns within a data stream is important for fields as diverse as cybersecurity or e-commerce. This requires to use pattern mining techniques. However, pattern mining suffers from two issues. The first one, known as “pattern explosion”, comes from the large combinatorial space explored and is the result of too many patterns output to be analyzed. Recent techniques called output space sampling solve this problem by outputting only a sampled set of all the results, with a target size provided by the user. The second issue is that most algorithms are designed to operate on static datasets or low throughput streams. In [9], we propose a contribution to tackle both issues, by designing an FPGA accelerator for pattern mining with output space sampling. We show that our accelerator can outperform a state-of-the-art implementation on a server class CPU using a modest FPGA product.

7.3. Statistically Significant Discriminative Patterns Searching

In [18], we propose a novel algorithm, named SSDPS, to discover patterns in two-class datasets. The SSDPS algorithm owes its efficiency to an original enumeration strategy of the patterns, which allows to exploit some degrees of anti-monotonicity on the measures of discriminance and statistical significance. Experimental results demonstrate that the performance of the SSDPS algorithm is better than others. In addition, the number of generated patterns is much less than the number of the other algorithms. Experiment on real data also shows that SSDPS effciently detects multiple SNPs combinations in genetic data.

7.4. Compressing and Querying Skypattern Cubes

Skypatterns are important since they enable to take into account user preference through Pareto-dominance. Given a set of measures, a skypattern query finds the patterns that are not dominated by others. In practice, different users may be interested in different measures, and issue queries on any subset of measures (a.k.a subspace). This issue was recently addressed by introducing the concept of skypattern cubes. However, such a structure presents high redundancy and is not well adapted for updating operations like adding or removing measures, due to the high costs of subspace computations in retrieving skypatterns. In [11], we propose a new structure called Compressed Skypattern Cube (abbreviated CSKYC), which concisely represents a skypattern cube, and gives an efficient algorithm to compute it. We thoroughly explore its properties and provide an efficient query processing algorithm. Experimental results show that our proposal allows to construct and to query a CSKYC very efficiently.

7.5. Semantics of Negative Sequential Patterns

In the field of pattern mining, a negative sequential pattern expresses behavior by a sequence of present and absent events. In [14], we shed light on the ambiguity of this notation and identify eight possible semantics with the relation of inclusion of a motif in a sequence. These semantics are illustrated and we are studying them formally. We thus propose dominance and equivalence relationships between these semantics, and we highlight new properties of anti-monotony. These results could be used to develop new efficient algorithms for mining frequent negative sequential patterns.

7.6. Admissible Generalizations of Examples as Rules

Rule learning is a data analysis task that consists in extracting rules that generalize examples. This is achieved by a plethora of algorithms. Some generalizations make more sense for the data scientists, called here admissible generalizations. The purpose of our work in [8] is to show formal properties of admissible generalizations. A formalization for generalization of examples is proposed allowing the expression of rule admissibility. Some admissible generalizations are captured by preclosure and capping operators. Also, we are interested in selecting supersets of examples that induce such operators. We then define classes of selection functions. This formalization is more particularly developed for examples with numerical attributes. Classes of such functions are associated with notions of generalization and they are used to comment some results of the CN2 algorithm [22].

7.7. Towards a Framework for Seasonal Time Series Forecasting Using Clustering

Seasonal behaviours are widely encountered in various applications. For instance, requests on web servers are highly influenced by our daily activities. Seasonal forecasting consists in forecasting the whole next season for a given seasonal time series. It may help a service provider to provision correctly the potentially required resources, avoiding critical situations of over or under provision. In [10], we propose a generic framework to make seasonal time series forecasting. The framework combines machine learning techniques 1) to identify the typical seasons and 2) to forecast the likelihood of having a season type in one season ahead. We study this framework by comparing the mean squared errors of forecasts for various settings and various datasets. The best setting is then compared to state-of-the-art time series forecasting methods. We show that it is competitive with then.

7.8. Towards Sustainable Dairy Management - A Machine Learning Enhanced Method for Estrus Detection

Our research tackles the challenge of milk production resource use efficiency in dairy farms with machine learning methods. Reproduction is a key factor for dairy farm performance since cows milk production begin with the birth of a calf. Therefore, detecting estrus, the only period when the cow is susceptible to pregnancy, is crucial for farm efficiency. Our goal is to enhance estrus detection (performance, interpretability), especially on the currently undetected silent estrus (35% of total estrus), and allow farmers to rely on automatic estrus detection solutions based on affordable data (activity, temperature). In [17] we first propose a novel approach with real-world data analysis to address both behavioral and silent estrus detection through machine learning methods. Second, we present LCE, a local cascade based algorithm that significantly outperforms a typical commercial solution for estrus detection, driven by its ability to detect silent estrus. Then, our study reveals the pivotal role of activity sensors deployment in estrus detection. Finally, we propose an approach relying on global and local (behavioral versus silent) algorithm interpretability (SHAP) to reduce the mistrust in estrus detection solutions.

7.9. A Distributed Multi-Sensor Machine Learning Approach to Earthquake Early Warning

Our research [21] aims to improve the accuracy of Earthquake Early Warning (EEW) systems by means of machine learning. EEW systems are designed to detect and characterize medium and large earthquakes before their damaging effects reach a certain location. Traditional EEW methods based on seismometers fail to accurately identify large earthquakes due to their sensitivity to the ground motion velocity. The recently introduced high-precision GPS stations, on the other hand, are ineffective to identify medium earthquakes due to its propensity to produce noisy data. In addition, GPS stations and seismometers may be deployed in large numbers across different locations and may produce a significant volume of data consequently, affecting the response time and the robust-ness of EEW systems. In practice, EEW can be seen as a typical classification problem in the machine learning field: multi-sensor data are given in input, and earthquake severity is the classification result. In this paper, we introduce the Distributed Multi-Sensor Earthquake Early Warning (DMSEEW) system, a novel machine learning-based approach that combines data from both types of sensors (GPS stations and seismometers) to detect medium and large earthquakes. DMSEEW is based on a new stacking ensemble method which has been evaluated on a real-world dataset validated with geoscientists. The system builds on a geographically distributed infrastructure, ensuring an efficient computation in terms of response time and robustness to partial infrastructure failures. Our experiments show that DMSEEW is more accurate than the traditional seismometer-only approach and the combined-sensors (GPS and seis-mometers) approach that adopts the rule of relative strength.

7.10. Dynamic Modeling of Nutrient Use and Individual Requirements of Lactating Sows

Nutrient requirements of sows during lactation are related mainly to their milk yield and feed intake, and vary greatly among individuals. In practice, nutrient requirements are generally determined at the population level based on average performance. The objective of the present modeling approach was to explore the variability in nutrient requirements among sows by combining current knowledge about nutrient use with on-farm data available on sows at farrowing [parity, BW, backfat thickness (BT)] and their individual performance (litter size, litter average daily gain, daily sow feed intake) to estimate nutrient requirements. The approach was tested on a database of 1,450 lactations from 2 farms. The effects of farm (A, B), week of lactation (W1: week 1, W2: week 2, W3+: week 3 and beyond), and parity (P1: 1, P2: 2, P3+: 3 and beyond) on sow performance and their nutrient requirements were evaluated. The mean daily ME requirement was strongly correlated with litter growth ($R2 = 0.95$; $P < 0.001$) and varied slightly according to sow BW, which influenced the maintenance cost. The mean daily standardized ileal digestible (SID) lysine requirement was influenced by farm, week of lactation, and parity. Variability in SID lysine requirement per kg feed was related mainly to feed intake ($R2 = 0.51$; $P < 0.001$) and, to a smaller extent, litter growth ($R2 = 0.27$; $P < 0.001$). It was lowest in W1 (7.0 g/kg), greatest in W2 (7.9 g/kg), and intermediate in W3+ (7.5 g/kg; $P < 0.001$) because milk production increased faster than feed intake capacity did. It was lower for P3+ (6.7 g/kg) and P2 sows (7.3 g/kg) than P1 sows (8.3 g/kg) due to the greater feed intake of multiparous sows. The SID lysine requirement per kg of feed was met for 80% of sows when supplies were 112 and 120% of the mean population requirement on farm A and B, respectively, indicating higher variability in requirements on farm B. Other amino acid and mineral requirements were influenced in the same way as SID lysine. In [5], we present a modeling approach that allows us to capture individual variability in the performance of sows and litters according to farm, stage of lactation, and parity. It is an initial step in the development of new types of models able to process historical farm data (e.g., for ex post assessment of nutrient requirements) and real-time data (e.g., to control precision feeding).

7.11. Temporal Models of Care Sequences for the Exploration of Medico-administrative Data

Pharmaco-epidemiology with medico-administrative databases enables the study of the impact of health products in real-life settings. These studies require to manipulate the raw data and the care trajectories, in order to identify pieces of data that may witness the medical information that is looked for. The manipulation can be seen as a querying process in which a query is a description of a medical pattern (e.g. occurrence of illness) with the available raw features from care trajectories (e.g. occurrence of medical procedures, drug deliveries, etc.). The more expressive is the querying process, the more accurate is the medical pattern search. The temporal dimension of care trajectories is a potential information that may improve the description of medical patterns. The objective of this work [6] is to propose a formal framework that would design a well-founded tool for querying care trajectories with temporal medical patterns. In this preliminary work, we present the problematic and we introduce a use case that illustrates the comparison of several querying formalisms.

7.12. Improving Domain Adaptation By Source Selection

Domain adaptation consists in learning from a source data distribution a model that will be used on a different target data distribution. The domain adaptation procedure is usually unsuccessful if the source domain is too different from the target one. In [16], we study domain adaptation for image classification with deep learning in the context of multiple available source domains. This work proposes a multi-source domain adaptation method that selects and weights the sources based on inter-domain distances. We provide encouraging results on both classical benchmarks and a new real world application with 21 domains.

7.13. From Cost-Sensitive Classification to Tight F-measure Bounds

The F-measure is a classification performance measure, especially suited when dealing with imbalanced datasets, which provides a compromise between the precision and the recall of a classifier. As this measure is non-convex and non-linear, it is often indirectly optimized using cost-sensitive learning (that affects different costs to false positives and false negatives). In [7], we derive theoretical guarantees that give tight bounds on the best F-measure that can be obtained from cost-sensitive learning. We also give an original geometric interpretation of the bounds that serves as an inspiration for CONE, a new algorithm to optimize for the F-measure. Using 10 datasets exhibiting varied class imbalance, we illustrate that our bounds are much tighter than previous work and show that CONE learns models with either superior F-measures than existing methods or comparable but in fewer iterations.

7.14. Time Series Classification Based on Interpretable Shapelets

[19] proposes a new architecture, called AI⟷PR-CNN, composed of generative adversarial neural networks (GANs), which addresses the problem of the lack of interpretability of the existing methods for time series classification. Our network has two components: a classifier and a discriminator. The classifier is a CNN, it serves to classify series. Convolutions are discriminant patterns learned from the data that allow for a more discriminating representation of time series (similar to a shapelet). To be able to explain the decision of the classifier, we would like to impose that the convolutions used are real “shapelets”, that is to say that they are close to real sub-series present in the training set. This constraint is implemented by a GAN whose purpose will determine how much the weight matrices classifier convolutions are close to subset of the training set.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

- **AdvisorSLA 2018 - Inria**
  Participants: E. Bourrand, L. Galárraga, E. Fromont, A. Termier
  Contract amount: 7,5k€
  Context. AdvisorSLA is a French company headquartered in Cesson-Sévigné, a city located in the outskirts of Rennes in Brittany. The company is specialized in software solutions for network

monitoring. For this purpose, the company relies on techniques of network metrology. AdvisorSLA's customers are carriers and telecommunications/data service providers that require to monitor the performance of their communication infrastructure as well as their QoE (quality of service). Network monitoring is of tremendous value for service providers because it is their primary tool for proper network maintenance. By continuously measuring the state of the network, monitoring solutions detect events (e.g., an overloaded router) that may degrade the network's operation and the quality of the services running on top of it (e.g., video transmission could become choppy). When a monitoring solution detects a potentially problematic sequence of events, it triggers an alarm so that the network manager can take actions. Those actions can be preventive or corrective. Some statistics gathered by the company show that only 40% of the triggered alarms are conclusive, that is, they manage to signal a well-understood problem that requires an action from the network manager. This means that the remaining 60% are presumably false alarms. While false alarms do not hinder network operation, they do incur an important cost in terms of human resources.
Objective. We propose to characterize conclusive and false alarms. This will be achieved by designing automatic methods to "learn" the conditions that most likely precede the fire of each type of alarm, and therefore predict whether the alarm will be conclusive or not. This can help adjust existing monitoring solutions in order to improve their accuracy. Besides, it can help network managers automatically trace the causes of a problem in the network. The aforementioned problem has an inherent temporal nature: we need to learn which events occur before an alarm and in which order. Moreover, metrology models take into account the measurements of different components and variables of the network such as latency and packet loss. For these two reasons, we resort to the field of multivariate time sequences and time series. The fact that we know the "symptoms" of an alarm and whether it is conclusive or not, allows for the application of supervised machine learning and pattern mining methods.
Additional remarks. This is a pre-doctoral contract signed with AdvisorSLA to start the work for the PhD of E. Bourrand (Thèse CIFRE) while the corresponding administrative formalities are completed.

- **ATERMES 2018-2021 - Univ Rennes 1**
  Participants: H. Zhang, E. Fromont
  Contract amount: 45k€
  Context. ATERMES is an international mid-sized company, based in Montigny-le-Bretonneux with a strong expertise in high technology and system integration from the upstream design to the long-life maintenance cycle. It has recently developed a new product, called BARIERTM ("Beacon Autonomous Reconnaissance Identification and Evaluation Response"), which provides operational and tactical solutions for mastering borders and areas. Once in place, the system allows for a continuous night and day surveillance mission with a small crew in the most unexpected rugged terrain. BARIER™ is expected to find ready application for temporary strategic site protection or ill-defined border regions in mountainous or remote terrain where fixed surveillance modes are impracticable or overly expensive to deploy.
  Objective. The project aims at providing a deep learning architecture and algorithms able to detect anomalies (mainly the presence of people or animals) from multimodal data. The data are considered "multimodal" because information about the same phenomenon can be acquired from different types of detectors, at different conditions, in multiple experiments, etc. Among possible sources of data available, ATERMES provides Doppler Radar, active-pixel sensor data (CMOS), different kind of infra-red data, the border context etc. The problem can be either supervised (if label of objects to detect are provided) or unsupervised (if only times series coming from the different sensors are available). Both the multimodal aspect and the anomaly detection one are difficult but interesting topics for which there exist few available works (that take both into account) in deep learning.

- **PSA - Inria**
  Participants: E. Fromont, A. Termier, L. Rozé, G. Martin
  Contract amount: 15k€
  Context. Peugeot-Citroën (PSA) group aims at improving the management of its car sharing service. To optimize its fleet and the availability of the cars throughout the city, PSA needs to analyze the trajectory of its cars.
  Objective. The aim of the internship is (1) to survey the existing methods to tackle the aforementioned need faced by PSA and (2) to also investigate how the techniques developed in LACODAM (e.g., emerging pattern mining) could be serve this purpose. A framework, consisting of three main modules, has been developped. We describe the modules in the following.
  - A town modelisation module with clustering. Similar towns are clustered in order to reuse information from one town in other towns.
  - A travel prediction module with basic statistics.
  - A reallocation strategy module (choices on how to relocate cars so that the most requested areas are always served). The aim of this module is to be able to test different strategies.

  Additional remarks. This is a pre-doctoral contract to start the work for the PhD of G. Martin (Thèse CIFRE) while the corresponding administrative formalities are completed.

9. Partnerships and Cooperations

9.1. National Initiatives

- **HyAIAI: Hybrid Approaches for Interpretable AI**
  Participants: E. Fromont (leader), A. Termier, L. Galárraga
  The Inria Project Lab HyAIAI is a consortium of Inria teams (Sequel, Magnet, Tau, Orpailleur, Multispeech, and LACODAM) that work together towards the development of novel methods for machine learning, that combine numerical and symbolic approaches. The goal is to develop new machine learning algorithms such that (i) they are as efficient as current best approaches, (ii) they can be guided by means of human-understandable constraints, and (iii) its decisions can be better understood.
- **Hyptser: Hybrid Prediction of Time Series**
  Participants: T. Guyet, S. Malinowski (LinkMedia), V. Lemaire (Orange)
  HYPTSER is a collaborative project between Orange Labs and LACODAM funded by the Fondation Mathématique Jacques Hadamard (PGMO program). It aims at developping new hybrid time series prediction methods in order to improve capacity planning for server farms. Capacity planning is the process of determining the infrastructure needed to meet future customer demands for online services. A well-made capacity planning helps to reduce operational costs, and improves the quality of the provided services. Capacity planning requires accurate forecasts of the differences between the customer demands and the infrastructure theoretical capabilities. The HYPTSER project makes the assumption that this information is captured by key performance indicators (KPI), that are measured continuously in the service infrastructure. Thus, we expect to improve capacity planning capabilities by making accurate forecasts of KPI time series. Recent methods about time series forecasting make use of ensemble models. In this project, we are interested in developing hybrid models for time series forecasting. Hybrid models aim at jointly partitioning the data, learning forecasting models in each partition and learning how to combine their outputs. We are currently developing two different approaches for that purpose, one based on the MODL framework and the other based on neural networks. We describe these approaches below:

  - MODL is a mathematical framework that turns the learning task into a model selection problem. It aims at finding the most probable model given the data. The MODL approach has been applied on numerous learning tasks. In all cases, this approach leads to a regularized optimization criterion. We formalize a new MODL criterion able to learn hybrid models on time series in order to: i) make a partition of time series; ii) learn local regression models. This approach formalizes these two steps in a unified way.
  - We are also developing an hybrid neural network structure that is able to learn automatically a soft partitioning of the data together with local models on each partition.

  In the next steps of this project, we will analyze the performance of this two strategies on KPI time series provided by Orange and compare them to classical ensemble methods.

- **#DigitAg: Digital Agriculture**
  Participants: A. Termier, V. Masson, C. Largouët, A.I. 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, which 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).

9.1.1. ANR

- **FAbLe: Framework for Automatic Interpretability in Machine Learning**
  Participants: L. Galárraga (holder), C. Largouët
  How can we fully automatically choose the best explanation for a given use case in classification?. Answering this question is the raison d'être of the JCJC ANR project FAbLe. By "best explanation" we mean the explanation that yields the best trade-off between interpretability and fidelity among a universe of possible explanations. While fidelity is well-defined as the accuracy of the explanation w.r.t the answers of the black-box, interpretability is a subjective concept that has not been formalized yet. Hence, in order to answer our prime question we first need to answer the question: "How can we formalize and quantify interpretability across models?". Much like research in automatic machine learning has delegated the task of accurate model selection to computers [26], FAbLe aims at fully delegating the selection of interpretable explanations to computers. Our goal is to produce a suite of algorithms that will compute suitable explanations for ML algorithms based on our insights of what is interpretable. The algorithms will choose the best explanation method based on the data, the use case, and the user's background. We will implement our algorithms so that they are fully compatible with the body of available software for data science (e.g., Scikit-learn).

9.1.2. National Platforms

- **PEPS: Pharmaco-epidemiology for Health Products**
  Participants: J. Bakalara, Y. Dauxais, T. Guyet, V. Masson, R. Quinou, A. 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. Since last year we have been working on the design and development of algorithms [25], [24] to mine patterns on care pathways.

9.2. International Research Visitors

9.2.1. Internships

From September to December 2019 we hosted Vaishnavi Bhargava, a computer science student from the Birla Institute of Technology and Science in Pilani, who worked on “Automatic Neighborhood Design for Localized Model-interpretation”. Her work aimed at finding a set of metrics and procedures to determine the best parameterization of the method LIME for local post-hoc interpretability of machine learning models. The goal of this effort is to inform users of the parameter values (if any) for which a LIME explanation should be trusted because it can faithfully reproduce the behavior of the black-box it tries to explain.

10. Dissemination

10.1. Scientific Events: Organisation

10.1.1. General Chair, Scientific Chair

- Elisa Fromont was one of the general chairs of ECML/PKDD 2019 that took place in Würzburg, Germany (Sept).
- Tassadit Bouadi was co-chair of the ECML/PKDD’2019 PhD Forum
- Tassadit Bouadi and Luis Galárraga are organizers of the two editions of the AIMLAI workshop (Advances in Interpretable Machine Learning and Artificial Intelligence) colocated with the conferences EGC (https://project.inria.fr/aimlai/) and ECML/PKDD (https://kdd.isti.cnr.it/xkdd2019/) 2019 respectively.
- AALTD@ECML (T. Guyet), https://project.inria.fr/aaltd19/
- GAST@EGC (T. Guyet), https://gt-gast.irisa.fr/gast-2019/
- GeoInformation Analytics Technical Track at the ACM/SIGAPP Symposium On Applied Computing (T. Guyet), https://gia.sciencesconf.org/
- Time series days, https://project.inria.fr/tsdays/ (T. Guyet)

10.1.2. Scientific Events: Selection

10.1.2.1. Member of the Conference Program Committees

- KDD (A. Termier, E. Fromont)
- ICTAI (T. Guyet)
- ECAI (T. Bouadi)
- ECML-PKDD (T. Guyet, A. Termier, E. Fromont, T. Bouadi)
- IJCAI (T. Guyet, A. Termier)
- SDM (A. Termier)
- MedInfo (T. Guyet)
- EGC (T. Bouadi, T. Guyet)
- HiPC (A. Termier)
- APIA (C. Largouët)
- CAp (R. Gaudel)
- ISWC (L. Galárraga)
- LDK (L. Galárraga)

10.1.2.2. Reviewer

- ECML-PKDD (C. Largouët)
- ICML (R. Gaudel)
- NeurIPS (R. Gaudel)
- LOD (R. Gaudel)
- ICTAI (L. Galárraga)
- BDA (L. Galárraga)

10.1.3. Journal

10.1.3.1. Reviewer - Reviewing Activities

- Pattern Recognition (R. Gaudel)
- Revue d'intelligence artificielle (R. Gaudel)
- Data Mining and Knowledge Discovery (L. Galárraga, A. Termier)
- Semantic Web Journal (L. Galárraga)
- Remote Sensing (T. Guyet)
- Journal of Biomedical Informatics (T. Guyet)

10.1.4. Invited Talks

- Signatures: detecting and characterizing recurrent behavior in sequential data (A. Termier, March 2019). Invited talk to the Exploratory Data Analysis group of Jilles Vreeken, Saarebruck University.
- Pattern Mining with MDL (A. Termier, May 2019). Talk at the Dagstuhl seminal on Enumeration in Data Management.
- (Random) Search in a Structured Space (R.Gaudel, Oct. 2019). GdR IA course, Paris, France.
- Recommendation as a sequential process (R. Gaudel, June 2019). Obelix, Vannes, France.
- Litterature review: From computational complexity of function optimization to large scale Machine Learning (R. Gaudel, June 2019). Thematic season on the topic COMPLEXITY, IRISA, Rennes, France
- Interpretability in Classifiers (L. Galárraga, May 2019). Journée EGC&IA, Orsay, France.
- Mining sequential patterns with ASP (T. Guyet, January 2019). LERIA, Angers, France. 6/12/2019: Master Class at La Digital Tech Conference on "IA, Machine learning et Deep learning : démêlons les concepts !", Rennes.
- (E. Fromont) 25/11/2019: Participation in the panel organized by EIT Digital for the "Industrial Doctorate meetup: Boosting business with data science", Rennes.
- (E. Fromont) 8/10/2019: Invited speaker for Planète conférence on "Mythes et réalités de l'I.A.", Vannes.
- (E. Fromont) 9/09/2019: Participation in the workshop "Smart-cities et mobilités" of the Forum Européen d'Intelligence Artificielle Territoriale organised by the CHEMI, Rambouillet, Fr.
- (E. Fromont) 5/07/2019: Participation in the panel organized by Google numérique on "Intelligence artificielle, opportunités diverses et défis communs?", Rennes.
- (E. Fromont) 1/07/2019: Participation in the panel organized by EIT summer school "Unleashing the power of data for better cities" on "Future of smart cities and data innovation", Rennes.
- (E. Fromont) 6/06/2019: Invited speaker for Breizh Conseil on "Mythes et réalités de l'I.A.", Rennes (with a video teaser)
- (E. Fromont) 5/06/2019: Invited speaker for the "journées nationales MIAGE" (JNM2019) on "Mythes et réalités de l'I.A.", Rennes.

- (E. Fromont) 23/05/2019: Invited speaker for ESTIM NUMERIQUE on "Qu'est ce que l'I.A.", Rennes.
- (E. Fromont) 31/01/2019: Invited speaker for the Ecole Navale Science Day 2019 JSN'19 on "New Challenges in Computer Vision", Brest.
- (E. Fromont) 26/01/2019: Invited speaker for the Digital transformer Challenge (Introduction to AI), Rennes.

10.1.5. Scientific Expertise

Scikit-Mine. The team recently secured a two-year engineer position (Inria ADT grant) for working on MDL-based pattern mining algorithms. The goal of the project is to design a modern pattern mining library fully compatible with scikit-learn (the major data science library). The library will contain major MDL-based pattern mining algorithms of the state of the art, including those developed in the team (periodic pattern mining). The project will start in fall 2020.

10.1.6. Research Administration

- Alexandre Termier is member of the Scientific Committee of the "Environment and Agriculture" department of INRA (now INRAe).
- (E. Fromont) Member of an HCERES evaluation committee.
- (E. Fromont) Member of the executive board of the CominLabs LABEX for the "Data, IA and Robotics" scientific axis.
- (E. Fromont) Elected at the scientific council of the Société Savante Francophone d'Apprentissage Machine.
- (E. Fromont) Nominated at the CNRS INS2I ("INstitut des Sciences de l'Information et de leurs Interactions") scientific council (CSI)

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Some 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 teachings LACODAM is involved in the following programs:

Master 2 IL, CCN: Option Machine Learning, Istic, University of Rennes 1, 32h (E. Fromont)

Master 2 DMV Module: Data Mining and Visualization, 13h, M2, Istic, University of Rennes 1 (A. Termier)

Master 2 Big Data, 30h, Master Datascience, University of Rennes 2 & Agrocampus Ouest (C. Largouët)

Master 2 DataViz with R, 10h, Master Datascience, University of Rennes 2 & Agrocampus Ouest (L. Bonneau de Beaufort)

Master 1 SIF: Option IA, Istic, 20h, University of Rennes 1 (A. Termier, E. Fromont, T. Bouadi)

Master 1 Scientific Programming, 25h, Agrocampus Ouest (C. Largouët)

Master 1 Data Management, 25h, Agrocampus Ouest (C. Largouët)

Master : R. Gaudel, Apprentissage profond, 9h eq. TD, M2, ENSAI, France

Master : R. Gaudel, Neural Networks, 4,5h eq. TD, école d'été, ENSAI, France

Master : R. Gaudel, Systèmes de recommandation, 27h eq. TD, M2, ENSAI, France

Master : R. Gaudel, Bandits Theory, 9h eq. TD, M2, ENSAI, France

Master : R. Gaudel, Bandits Theory, 4,5h eq. TD, école d'été, ENSAI, France

Master : R. Gaudel, INFO2, 30h eq. TD, M1, ENS, France

Master : T. Guyet, Spatial data science and GIS programming, Agrocampus-Ouest/Rennes University, France

10.2.2. Supervision

PhD (defended 25/09/2019): Guillaume Metzler, “Learning from Imbalanced Data: An Application to Bank Fraud Detection”, 15/01/2016, E. Fromont, M. Sebban, A. Habrard.

PhD (defended 16/12/2019): Kevin Bascol, “Multi-source domain adaptation on imbalanced data:application to the improvement of chairlifts safety”, 1/11/2016, E. Fromont,R. Emonet.

PhD (defended 16/12/2019): Maël Guillemé, “New Data Mining Approaches for Improving Energy Consumption in Factories”, 03/10/2016, L. Rozé, V. Masson, A. Termier

PhD (defended 19/12/2019): Alban Siffer, “Data Mining Approaches for Cyber Attack Detection”, 03/2016, P-A Fouque, A. Termier, C. Largouët.

PhD in progress: Maël Gueguen, “Improving the Performance and Energy Efficiency of Complex Heterogeneous Manycore Architectures with On-chip Data Mining”, 01/11/2016, O. Sentieys, A. Termier

PhD in progress: Raphaël Gauthier, “Modelling of Nutrient Utilization and Precision Feeding of Lactating Sows”, 01/11/2017, C. Largouët, J.-Y. Dourmad

PhD in progress: Kévin Fauvel, “Using Data mining Techniques for Improving Dairy Management”, 01/10/2017, V. Masson, A. Termier, P. Faverdin

PhD in progress: Heng Zhang, “Deep Learning on Multimodal Data for the Supervision of Sensitive Sites”, 03/12/2018, E. Fromont, S. Lefevre

PhD in progress: Yichang Wang, “Interpretable Shapelet for Anomaly Detection in Time Series”, 15/04/2018, E. Fromont, S. Malinowski, R. Tavenard, R. Emonet

PhD in progress: Camille-Sauvanneary Gauthier, Session-aware Recommendation System, since March 15. 2019, advisors: É. Fromont, R. Gaudel, & B. Guilbot

PhD in progress: Erwan Bourrand, Interactive Data Mining for Root Cause Analysis of Performance Issues in Networks, advisors: E. Fromont, L. Galárraga, A. Termier.

PhD in progress: Gregory Martin, Data mining to optimize a free-floating car sharing service: E. Fromont, L. Rozé, A. Termier.

PhD in progress: Johanne Bakalara, Temporal models of care sequences for the exploration of medico-administrative data, advisors: E. Oger, O. Dameron, T. Guyet, A. Happe

PhD in progress: Colin Leverger, Management and forecasting of measures for server optimization in a cloud context, advisors: A. Termier, R. Marguerie, S. Malinowski, T. Guyet, L. Rozé.

PhD in progress: Nicolas Sourbier, Log Analysis Using Reinforcement Learning for Intrusion Detection, advisors: F. Majorczyke, O. Gesny, T. Ollivier, M. Pelcat, T. Guyet.

PhD in progress: Yiru Zhang, Modeling and management of imperfect preferences with the theory of belief functions, advisors: T. Bouadi, A. Martin.

10.2.3. Juries

- PhD Justine Reynaud (Orpailleur, Inria Nancy, Université de Lorraine): Découverte de définitions dans le web des données (L. Galárraga, examiner, december 2019).
- PhD Anes Bendimerad (INSA Lyon): Mining useful patterns in attributed graphs (A. Termier, reviewer, september 2019)
- HDR Frédéric Flouvat (Université de Nouvelle Calédonie): Extraction de motifs spatio-temporels: co-localisations, séquences et graphes dynamiques attribués (A. Termier, reviewer, october 2019)
- HDR Christine Largouët (Université de Rennes 1): Intelligence Artificielle pour l’aide à la décision des systèmes dynamiques: diagnostic, prévision, recommandation d’actions (A. Termier, examiner, december 2019)
- PhD Ricardio Sperandio (Université de Rennes 1): Time series retrieval using DTW-preserving shapelets (A. Termier, examiner, december 2019)

- PhD Raphaël Jakse (Université Grenoble Alpes): Interactive runtime verification (A. Termier, reviewer, december 2019)
- (E. Fromont) Ziyu GUO, Marseille 5/11/2019 (committee member); Jordan Frery, Saint-Etienne 26/09/2019 (committee member, president); Julien Salotti, Lyon 24/09/2019 (reviewer); Xuhong Li, Compiègne (committee member, president) 10/99/2019; Michele Linardi, Paris (committee member) 21/08/2019; Pauline Luc, Grenoble (committee member, president) 25/06/2019, Diogo Luvizon, Cergy (reviewer) 8/04/2019; Zaruhi Alaverdyan, Lyon (committee member, president) 18/01/2019
- PhD Trung-Dung Le (Shaman, IRISA, Université de Rennes 1): Gestion de masses de données dans une fédération de nuages informatiques (T. Bouadi, examiner, july 2019)

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

- (E. Fromont) 2018- ?: Nominated at the scientific council of the "Fondation Blaise Pascal" (dedicated to scientific mediation, funded by Inria)

10.3.2. Education

- In 2019, Alexandre Termier and Elisa Fromont are co-heads of a special training for high school teachers, ordered by the Ministry of Education. From September 2019, French high school students will have the opportunity to have reinforced Computer Science courses, with the same importance as Mathematics or Physics for example. As new teachers are not yet hired to teach theses courses, there was thus an urgent need to train existing high school teachers so that they can train the new teachers. A national formation has been set up, with some coordination between all universities. Alexandre and Elisa represent the University of Rennes, which is the formation center for all of Britany. The formation load is 125 teaching hours, to a group of 71 high school teachers.
- Elisa Fromont is also leading the set up of a CAPES in University of Rennes 1, which will allow to prepare future high school teachers in computer science.

10.3.3. Interventions

- A. Termier participated at a mediation action on Artificial Intelligence for a general technical public, the "Trial of metro car number 42", at the Digital Tech in Rennes (november 2019, https://www.ladigital.tech/proces-rame-metro-numero-42/)
- Alexandre Termier did two 2h interventions at the Emile Zola high school, aimed at math teachers. The topics were "Artificial Intelligence" and "Big Data".
- (E. Fromont) 6/12/2019: Master Class at La Digital Tech Conference on "IA, Machine learning et Deep learning : démêlons les concepts !", Rennes.

11. Bibliography

Major publications by the team in recent years

[1] K. FAUVEL, V. MASSON, E. FROMONT, P. FAVERDIN, A. TERMIER.*Towards Sustainable Dairy Management - A Machine Learning Enhanced Method for Estrus Detection*, in "KDD 2019 - ACM SIGKDD International Conference on Knowledge Discovery & Data Mining", Anchorage, United States, 25th SIGKDD Conference on Knowledge Discovery and Data Mining proceedings, August 2019, p. 1-9 [*DOI :* 10.1145/3292500.3330712], https://hal.archives-ouvertes.fr/hal-02190790

[2] E. GALBRUN, P. CELLIER, N. TATTI, A. TERMIER, B. CRÉMILLEUX.*Mining Periodic Patterns with a MDL Criterion*, in "European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD)", Dublin, Ireland, 2018, https://hal.archives-ouvertes.fr/hal-01951722

Publications of the year

Doctoral Dissertations and Habilitation Theses

[3] C. LARGOUËT.*Artificial Intelligence for decision support in dynamic systems: Diagnosis, Prediction, Recommendation of actions*, Université de Rennes 1, December 2019, Habilitation à diriger des recherches, https://hal.inria.fr/tel-02437159

Articles in International Peer-Reviewed Journal

[4] A. AYADI, A. SAMET, F. D. B. DE BEUVRON, C. ZANNI-MERK.*Ontology population with deep learning-based NLP: a case study on the Biomolecular Network Ontology*, in "Procedia Computer Science", 2019, vol. 159, p. 572-581 [*DOI :* 10.1016/J.PROCS.2019.09.212], https://hal.archives-ouvertes.fr/hal-02357839

[5] R. GAUTHIER, C. LARGOUËT, C. GAILLARD, L. CLOUTIER, F. GUAY, J.-Y. DOURMAD.*Dynamic modeling of nutrient use and individual requirements of lactating sows*, in "Journal of Animal Science", May 2019, vol. 97, n^o 7, p. 2822–2836 [*DOI :* 10.1093/JAS/SKZ167], https://hal.archives-ouvertes.fr/hal-02142870

International Conferences with Proceedings

[6] J. BAKALARA, T. GUYET, O. DAMERON, E. OGER, A. HAPPE.*Temporal models of care sequences for the exploration of medico-administrative data*, in "2019 - Workshop IA&Santé, PFIA", Toulouse, France, July 2019, p. 1-8, https://hal.archives-ouvertes.fr/hal-02265743

[7] K. BASCOL, R. EMONET, E. FROMONT, A. HABRARD, G. METZLER, M. SEBBAN.*From Cost-Sensitive Classification to Tight F-measure Bounds*, in "AISTATS 2019 - 22nd International Conference on Artificial Intelligence and Statistics", Naha, Okinawa, Japan, The 22nd International Conference on Artificial Intelligence and Statistics, April 2019, vol. 89, n^o 1, p. 1245-1253, https://hal.archives-ouvertes.fr/hal-02049763

[8] P. BESNARD, T. GUYET, V. MASSON.*Admissible Generalizations of Examples as Rules*, in "ICTAI 2019 - 31st International Conference on Tools with Artificial Intelligence", Portland, United States, November 2019, p. 1480-1485, https://hal.inria.fr/hal-02267166

[9] M. GUEGUEN, O. SENTIEYS, A. TERMIER.*Accelerating Itemset Sampling using Satisfiability Constraints on FPGA*, in "DATE 2019 - 22nd IEEE/ACM Design, Automation and Test in Europe", Florence, Italy, IEEE, March 2019, p. 1046-1051 [*DOI :* 10.23919/DATE.2019.8714932], https://hal.inria.fr/hal-01941862

[10] C. LEVERGER, S. MALINOWSKI, T. GUYET, V. LEMAIRE, A. BONDU, A. TERMIER.*Toward a Framework for Seasonal Time Series Forecasting Using Clustering*, in "IDEAL 2019", Manchester, United Kingdom, October 2019, p. 328-340 [*DOI :* 10.1007/978-3-030-33607-3_36], https://hal.inria.fr/hal-02371221

[11] W. UGARTE, S. LOUDNI, P. BOIZUMAULT, B. CRÉMILLEUX, A. TERMIER.*Compressing and Querying Skypattern Cubes*, in "IEA/AIE-2019 - 32nd International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems", Graz, Austria, Advances and Trends in Artificial Intelligence. From Theory to Practice 32nd International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2019, Graz, Austria, July 9–11, 2019, Proceedings, Springer, July 2019, p. 406-421 [*DOI :* 10.1007/978-3-030-22999-3_36], https://hal.archives-ouvertes.fr/hal-02190788

National Conferences with Proceeding

[12] R. GAUTHIER, C. LARGOUËT, C. GAILLARD, L. CLOUTIER, F. GUAY, J.-Y. DOURMAD.*Dynamic modelling of nutrient use and individual requirements in lactating sows*, in "JRP 2019 - 51èmes Journées de la Recherche Porcine", Paris, France, IFIP - Institut du Porc, 2019, p. 117-122, https://hal.archives-ouvertes.fr/hal-02010788

[13] T. GUYET, P. BESNARD, A. SAMET, N. BEN SALHA, N. LACHICHE.*Énumération des occurrences d'une chronique*, in "EGC 2020 - 20ème édition de la Conférence Extraction et Gestion des Connaissances", Bruxelles, Belgium, January 2020, p. 1-8, https://hal.inria.fr/hal-02422796

[14] T. GUYET.*Semantic(s) of negative sequential patterns*, in "JIAF 2019 - 13èmes Journées d'Intelligence Artificielle Fondamentale", Toulouse, France, July 2019, p. 1-10, https://hal.inria.fr/hal-02188501

Conferences without Proceedings

[15] J. BAKALARA.*Temporal models of care sequences for the exploration of medico-administrative data*, in "AIME 2019 - 17th Conference on Artificial Intelligence in Medicine", Poznan, Poland, June 2019, p. 1-7, https://hal.archives-ouvertes.fr/hal-02265731

[16] K. BASCOL, R. EMONET, E. FROMONT.*Improving Domain Adaptation By Source Selection*, in "ICIP 2019 - IEEE International Conference on Image Processing", Taïpei, Taiwan, IEEE, September 2019 [*DOI :* 10.1109/ICIP.2019.8803325], https://hal.archives-ouvertes.fr/hal-02136378

[17] K. FAUVEL, V. MASSON, E. FROMONT, P. FAVERDIN, A. TERMIER.*Towards Sustainable Dairy Management - A Machine Learning Enhanced Method for Estrus Detection*, in "KDD 2019 - ACM SIGKDD International Conference on Knowledge Discovery & Data Mining", Anchorage, United States, 25th SIGKDD Conference on Knowledge Discovery and Data Mining proceedings, August 2019, p. 1-9 [*DOI :* 10.1145/3292500.3330712], https://hal.archives-ouvertes.fr/hal-02190790

[18] H. S. PHAM, G. VIRLET, D. LAVENIER, A. TERMIER.*Statistically Significant Discriminative Patterns Searching*, in "DaWaK 2019 - 21st International Conference on Big Data Analytics and Knowledge Discovery", Linz, Austria, Springer, August 2019, p. 105-115 [*DOI :* 10.1007/978-3-030-27520-4_8], https://hal.archives-ouvertes.fr/hal-02190793

[19] Y. WANG, R. EMONET, E. FROMONT, S. MALINOWSKI, E. MENAGER, L. MOSSER, R. TAVENARD.*Classification de séries temporelles basée sur des "shapelets" interprétables par réseaux de neurones antagonistes*, in "CAp 2019 - Conférence sur l'Apprentissage automatique", Toulouse, France, July 2019, p. 1-2, https://hal.archives-ouvertes.fr/hal-02268004

Scientific Books (or Scientific Book chapters)

[20] Y. DAUXAIS, D. GROSS-AMBLARD, T. GUYET, A. HAPPE.*Discriminant chronicle mining*, in "Advances in Knowledge Discovery and Management (vol 8)", B. PINAUD, F. GUILLET, F. GANDON, C. LARGERON (editors), Advances in Knowledge Discovery and Management, Springer, Cham, June 2019, p. 89–118 [*DOI :* 10.1007/978-3-030-18129-1_5], https://hal.inria.fr/hal-01940146

Other Publications

[21] K. FAUVEL, D. BALOUEK-THOMERT, D. MELGAR, P. SILVA, A. SIMONET, G. ANTONIU, A. COSTAN, V. MASSON, M. PARASHAR, I. RODERO, A. TERMIER.*A Distributed Multi-Sensor Machine Learning*

Approach to Earthquake Early Warning, November 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02373429

References in notes

[22] P. CLARK, T. NIBLETT.*The CN2 Induction Algorithm*, in "Mach. Learn.", March 1989, vol. 3, n^o 4, p. 261–283, https://doi.org/10.1023/A:1022641700528

[23] 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^o 10, p. 979–988

[24] 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

[25] 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 [*DOI :* 10.1007/978-3-319-59758-46], https://hal.archives-ouvertes.fr/hal-01568929

[26] M. FEURER, A. KLEIN, K. EGGENSPERGER, J. SPRINGENBERG, M. BLUM, F. HUTTER.*Efficient and Robust Automated Machine Learning*, in "Advances in Neural Information Processing Systems 28", C. CORTES, N. D. LAWRENCE, D. D. LEE, M. SUGIYAMA, R. GARNETT (editors), Curran Associates, Inc., 2015, p. 2962–2970, http://papers.nips.cc/paper/5872-efficient-and-robust-automated-machine-learning.pdf

[27] R. GUIDOTTI, A. MONREALE, S. RUGGIERI, F. TURINI, F. GIANNOTTI, D. PEDRESCHI.*A Survey of Methods for Explaining Black Box Models*, in "ACM Computing Surveys", 2018, vol. 51, n^o 5, p. 93:1–93:42

[28] T. KULESZA, S. STUMPF, M. M. BURNETT, S. YANG, I. KWAN, W.-K. WONG.*Too much, too little, or just right? Ways explanations impact end users' mental models*, in "2013 IEEE Symposium on Visual Languages and Human Centric Computing", 2013, p. 3-10

[29] 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^o 6, p. 411 - 417

[30] 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^o 49, e2229

[31] 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^o 11, p. 1161–1169

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

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
A5.7.1. - Sound
A5.7.3. - Speech
A5.8. - Natural language processing
A9.2. - Machine learning
A9.3. - Signal analysis
A9.4. - Natural language processing

Other Research Topics and Application Domains:

B9. - Society and Knowledge
B9.3. - Medias
B9.6.10. - Digital humanities
B9.10. - Privacy

1. Team, Visitors, External Collaborators

Research Scientists

Laurent Amsaleg [Team leader, CNRS, Senior Researcher, HDR]
Guillaume Gravier [CNRS, Senior Researcher, HDR]
Yannis Avrithis [Inria, Advanced Research Position]
Vincent Claveau [CNRS, Researcher]
Teddy Furon [Inria, Researcher, HDR]
Miaojing Shi [Inria, Starting Research Position, until Nov 2019]

Faculty Members

Ewa Kijak [Univ de Rennes I, Associate Professor]
Simon Malinowski [Univ de Rennes I, Associate Professor]
Pascale Sébillot [INSA Rennes, Professor, HDR]

Technical Staff

Mateusz Budnik [Inria, Engineer]
Jean Hany [Inria, Engineer, until Nov 2019]
Guillaume Le Noe-Bienvenu [CNRS, Engineer, from Nov 2019]
Cédric Maigrot [Inria, Engineer, from May 2019]
Florent Michel [Inria, Engineer, from May 2019]

PhD Students

Benoit Bonnet [Inria, from Nov 2019]

Ricardo Carlini Sperandio [Univ Federal de Minas Gerais, until Jul 2019]
Clément Dalloux [CNRS, until Nov 2019]
Mikail Demirdelen [Univ de Rennes I]
Cheikh Brahim El Vaigh [Inria]
Marzieh Gheisari Khorasgani [Inria]
Yann Lifchitz [Safran Tech, CIFRE]
Cyrielle Mallart [SIPA-Ouest France, CIFRE]
Oriane Siméoni [Inria]
François Torregrossa [Pages Jaunes, CIFRE]
Hanwei Zhang [Chinese Scholarship Council]

Post-Doctoral Fellow
Suresh Kirthi Kumaraswamy [Univ Le Mans, from Mar 2019]

Visiting Scientist
Raquel Pereira de Almeida [Univ Federal de Minas Gerais, from Feb 2019]

Administrative Assistant
Aurélie Patier [Univ de Rennes I]

2. Overall Objectives

2.1. Context

LINKMEDIA is concerned with the processing of extremely large collections of multimedia material. The material we refer to are collections of documents that are created by humans and intended for humans. It is material that is typically created by media players such as TV channels, radios, newspapers, archivists (BBC, INA, ...), as well as the multimedia material that goes through social-networks. It also includes material that includes images, videos and pathology reports for e-health applications, or that is in relation with e-learning which typically includes a fair amount of texts, graphics, images and videos associating in new ways teachers and students. It also includes material in relation with humanities that study societies through the multimedia material that has been produced across the centuries, from early books and paintings to the latest digitally native multimedia artifacts. Some other multimedia material are out of the scope of LINKMEDIA, such as the ones created by cameras or sensors in the broad areas of video-surveillance or satellite images.

Multimedia collections are rich in contents and potential, that richness being in part within the documents themselves, in part within the relationships between the documents, in part within what humans can discover and understand from the collections before materializing its potential into new applications, new services, new societal discoveries, ... That richness, however, remains today hardly accessible due to the conjunction of several factors originating from the inherent nature of the collections, the complexity of bridging the semantic gap or the current practices and the (limited) technology:

- *Multimodal:* multimedia collections are composed of very diverse material (images, texts, videos, audio, ...), which require sophisticated approaches at analysis time. Scientific contributions from past decades mostly focused on analyzing each media in isolation one from the other, using modality-specific algorithms. However, revealing the full richness of collections calls for jointly taking into account these multiple modalities, as they are obviously semantically connected. Furthermore, involving resources that are external to collections, such as knowledge bases, can only improve gaining insight into the collections. Knowledge bases form, in a way, another type of modality with specific characteristics that also need to be part of the analysis of media collections. Note that determining what a document is about possibly mobilizes a lot of resources, and this is especially costly and time consuming for audio and video. Multimodality is a great source of richness, but causes major difficulties for the algorithms running analysis;

- *Intertwined*: documents do not exist in isolation one from the other. There is more knowledge in a collection than carried by the sum of its individual documents and the relationships between documents also carry a lot of meaningful information. (Hyper)Links are a good support for materializing the relationships between documents, between parts of documents, and having analytic processes creating them automatically is challenging. Creating semantically rich typed links, linking elements at very different granularities is very hard to achieve. Furthermore, in addition to being disconnected, there is often no strong structure into each document, which makes even more difficult their analysis;
- *Collections are very large:* the scale of collections challenges any algorithm that runs analysis tasks, increasing the duration of the analysis processes, impacting quality as more irrelevant multimedia material gets in the way of relevant ones. Overall, scale challenges the complexity of algorithms as well as the quality of the result they produce;
- *Hard to visualize*: It is very difficult to facilitate humans getting insight on collections of multimedia documents because we hardly know how to display them due to their multimodal nature, or due to their number. We also do not know how to well present the complex relationships linking documents together: granularity matters here, as full documents can be linked with small parts from others. Furthermore, visualizing time-varying relationships is not straightforward. Data visualization for multimedia collections remains quite unexplored.

2.2. Scientific objectives

The ambition of LINKMEDIA is to propose **foundations, methods, techniques and tools to help humans make sense of extremely large collections of multimedia material**. Getting useful insight from multimedia is only possible if tools and users interact tightly. Accountability of the analysis processes is paramount in order to allow users understanding their outcome, to understand why some multimedia material was classified this way, why two fragments of documents are now linked. It is key for the acceptance of these tools, or for correcting errors that will exist. Interactions with users, facilitating analytics processes, taking into account the trust in the information and the possible adversarial behaviors are topics LINKMEDIA addresses.

3. Research Program

3.1. Scientific background

LINKMEDIA is de facto a multidisciplinary research team in order to gather the multiple skills needed to enable humans to gain insight into extremely large collections of multimedia material. It is *multimedia data* which is at the core of the team and which drives the design of our scientific contributions, backed-up with solid experimental validations. *Multimedia data*, again, is the rationale for selecting problems, applicative fields and partners.

Our activities therefore include studying the following scientific fields:

- multimedia: content-based analysis; multimodal processing and fusion; multimedia applications;
- computer vision: compact description of images; object and event detection;
- machine learning: deep architectures; structured learning; adversarial learning;
- natural language processing: topic segmentation; information extraction;
- information retrieval: high-dimensional indexing; approximate k-nn search; embeddings;
- data mining: time series mining; knowledge extraction.

3.2. Workplan

Overall, LINKMEDIA follows two main directions of research that are (i) extracting and representing information from the documents in collections, from the relationships between the documents and from what user build from these documents, and (ii) facilitating the access to documents and to the information that has been elaborated from their processing.

3.2.1. Research Direction 1: Extracting and Representing Information

LINKMEDIA follows several research tracks for *extracting* knowledge from the collections and *representing* that knowledge to facilitate users acquiring gradual, long term, constructive insights. Automatically processing documents makes it crucial to consider the accountability of the algorithms, as well as understanding when and why algorithms make errors, and possibly invent techniques that compensate or reduce the impact of errors. It also includes dealing with malicious adversaries carefully manipulating the data in order to compromise the whole knowledge extraction effort. In other words, LINKMEDIA also investigates various aspects related to the *security* of the algorithms analyzing multimedia material for knowledge extraction and representation.

Knowledge is not solely extracted by algorithms, but also by humans as they gradually get insight. This human knowledge can be materialized in computer-friendly formats, allowing algorithms to use this knowledge. For example, humans can create or update ontologies and knowledge bases that are in relation with a particular collection, they can manually label specific data samples to facilitate their disambiguation, they can manually correct errors, etc. In turn, knowledge provided by humans may help algorithms to then better process the data collections, which provides higher quality knowledge to humans, which in turn can provide some better feedback to the system, and so on. This virtuous cycle where algorithms and humans cooperate in order to make the most of multimedia collections requires specific support and techniques, as detailed below.

3.2.1.1. Machine Learning for Multimedia Material.

Many approaches are used to extract relevant information from multimedia material, ranging from very low-level to higher-level descriptions (classes, captions, ...). That diversity of information is produced by algorithms that have varying degrees of supervision. Lately, fully supervised approaches based on deep learning proved to outperform most older techniques. This is particularly true for the latest developments of Recurrent Neural Networkds (RNN, such as LSTMs) or convolutional neural network (CNNs) for images that reach excellent performance [62]. LINKMEDIA contributes to advancing the state of the art in computing representations for multimedia material by investigating the topics listed below. Some of them go beyond the very processing of multimedia material as they also question the fundamentals of machine learning procedures when applied to multimedia.

- *Learning from few samples/weak supervisions.* CNNs and RNNs need large collections of carefully annotated data. They are not fitted for analyzing datasets where few examples per category are available or only cheap image-level labels are provided. LINKMEDIA investigates low-shot, semi-supervised and weakly supervised learning processes: Augmenting scarce training data by automatically propagating labels [65], or transferring what was learned on few very well annotated samples to allow the precise processing of poorly annotated data [74]. Note that this context also applies to the processing of heritage collections (paintings, illuminated manuscripts, ...) that strongly differ from contemporary natural images. Not only annotations are scarce, but the learning processes must cope with material departing from what standard CNNs deal with, as classes such as "planes", "cars", etc, are irrelevant in this case.
- *Ubiquitous Training.* NN (CNNs, LSTMs) are mainstream for producing representations suited for high-quality classification. Their training phase is ubiquitous because the same representations can be used for tasks that go beyond classification, such as retrieval, few-shot, meta- and incremental learning, all boiling down to some form of metric learning. We demonstrated that this ubiquitous training is relatively simpler [65] yet as powerful as ad-hoc strategies fitting specific tasks [79]. We study the properties and the limitations of this ubiquitous training by casting metric learning as a classification problem.
- *Beyond static learning*. Multimedia collections are by nature continuously growing, and ML processes must adapt. It is not conceivable to re-train a full new model at every change, but rather to support continuous training and/or allowing categories to evolve as the time goes by. New classes may be defined from only very few samples, which links this need for dynamicity to the low-shot learning problem discussed here. Furthermore, active learning strategies determining which is the next sample to use to best improve classification must be considered to alleviate the annotation cost and the re-training process [69]. Eventually, the learning process may need to manage an extremely

large number of classes, up to millions. In this case, there is a unique opportunity of blending the expertise of LINKMEDIA on large scale indexing and retrieval with deep learning. Base classes can either be "summarized" e.g. as a multi-modal distribution, or their entire training set can be made accessible as an external associative memory [86].

- *Learning and lightweight architectures.* Multimedia is everywhere, it can be captured and processed on the mobile devices of users. It is necessary to study the design of lightweight ML architectures for mobile and embedded vision applications. Inspired by [90], we study the savings from quantizing hyper-parameters, pruning connections or other approximations, observing the trade-off between the footprint of the learning and the quality of the inference. Once strategy of choice is progressive learning which early aborts when confident enough [70].
- *Multimodal embeddings.* We pursue pioneering work of LINKMEDIA on multimodal embedding, i.e., representing multiple modalities or information sources in a single embedded space [83], [85], [84]. Two main directions are explored: exploiting adversarial architectures (GANs) for embedding via translation from one modality to another, extending initial work in [84] to highly heterogeneous content; combining and constraining word and RDF graph embeddings to facilitate entity linking and explanation of lexical co-occurrences [81].
- *Accountability of ML processes.* ML processes achieve excellent results but it is mandatory to verify that accuracy results from having determined an adequate problem representation, and not from being abused by artifacts in the data. LINKMEDIA designs procedures for at least explaining and possibly interpreting and understanding what the models have learned. We consider heat-maps materializing which input (pixels, words) have the most importance in the decisions [77], Taylor decompositions to observe the individual contributions of each relevance scores or estimating LID [47] as a surrogate for accounting for the smoothness of the space.
- *Extracting information.* ML is good at extracting features from multimedia material, facilitating subsequent classification, indexing, or mining procedures. LINKMEDIA designs extraction processes for identifying parts in the images [75], [76], relationships between the various objects that are represented in images [53], learning to localizing objects in images with only weak, image-level supervision [78] or fine-grained semantic information in texts [58]. One technique of choice is to rely on generative adversarial networks (GAN) for learning low-level representations. These representations can e.g. be based on the analysis of density [89], shading, albedo, depth, etc.
- *Learning representations for time evolving multimedia material.* Video and audio are time evolving material, and processing them requests to take their time line into account. In [71], [57] we demonstrated how shapelets can be used to transform time series into time-free high-dimensional vectors, preserving however similarities between time series. Representing time series in a metric space improves clustering, retrieval, indexing, metric learning, semi-supervised learning and many other machine learning related tasks. Research directions include adding localization information to the shapelets, fine-tuning them to best fit the task in which they are used as well as designing hierarchical representations.

3.2.1.2. Adversarial Machine Learning.

Systems based on ML take more and more decisions on our behalf, and maliciously influencing these decisions by crafting adversarial multimedia material is a potential source of dangers: a small amount of carefully crafted noise imperceptibly added to images corrupts classification and/or recognition. This can naturally impact the insight users get on the multimedia collection they work with, leading to taking erroneous decisions e.g.

This adversarial phenomenon is not particular to deep learning, and can be observed even when using other ML approaches [52]. Furthermore, it has been demonstrated that adversarial samples generalize very well across classifiers, architectures, training sets. The reasons explaining why such tiny content modifications succeed in producing severe errors are still not well understood.

We are left with little choice: we must gain a better understanding of the weaknesses of ML processes, and in particular of deep learning. We must understand why attacks are possible as well as discover mechanisms protecting ML against adversarial attacks (with a special emphasis on convolutional neural networks). Some initial contributions have started exploring such research directions, mainly focusing on images and computer vision problems. Very little has been done for understanding adversarial ML from a *multimedia* perspective [56].

LINKMEDIA is in a unique position to throw at this problem new perspectives, by experimenting with other modalities, used in isolation one another, as well as experimenting with true multimodal inputs. This is very challenging, and far more complicated and interesting than just observing adversarial ML from a computer vision perspective. No one clearly knows what is at stake with adversarial audio samples, adversarial video sequences, adversarial ASR, adversarial NLP, adversarial OCR, all this being often part of a sophisticated multimedia processing pipeline.

Our ambition is to lead the way for initiating investigations where the full diversity of modalities we are used to work with in multimedia are considered from a perspective of adversarial attacks and defenses, both at learning and test time. In addition to what is described above, and in order to trust the multimedia material we analyze and/or the algorithms that are at play, LINKMEDIA investigates the following topics:

- *Beyond classification.* Most contributions in relation with adversarial ML focus on classification tasks. We started investigating the impact of adversarial techniques on more diverse tasks such as retrieval [46]. This problem is related to the very nature of euclidean spaces where distances and neighborhoods can all be altered. Designing defensive mechanisms is a natural companion work.
- *Detecting false information.* We carry-on with earlier pioneering work of LINKMEDIA on false information detection in social media. Unlike traditional approaches in image forensics [60], we build on our expertise in content-based information retrieval to take advantage of the contextual information available in databases or on the web to identify out-of-context use of text or images which contributed to creating a false information [72].
- *Deep fakes.* Progress in deep ML and GANs allow systems to generate realistic images and are able to craft audio and video of existing people saying or doing things they never said or did [68]. Gaining in sophistication, these machine learning-based "deep fakes" will eventually be almost indistinguishable from real documents, making their detection/rebutting very hard. LINKMEDIA develops deep learning based counter-measures to identify such modern forgeries. We also carry on with making use of external data in a provenance filtering perspective [91] in order to debunk such deep fakes.
- *Distributions, frontiers, smoothness, outliers.* Many factors that can possibly explain the adversarial nature of some samples are in relation with their distribution in space which strongly differs from the distribution of natural, genuine, non adversarial samples. We are investigating the use of various information theoretical tools that facilitate observing distributions, how they differ, how far adversarial samples are from benign manifolds, how smooth is the feature space, etc. In addition, we are designing original adversarial attacks and develop detection and curating mechanisms [47].

3.2.1.3. Multimedia Knowledge Extraction.

Information obtained from collections via computer ran processes is not the only thing that needs to be represented. Humans are in the loop, and they gradually improve their level of understanding of the content and nature of the multimedia collection. Discovering knowledge and getting insight is involving multiple people across a long period of time, and what each understands, concludes and discovers must be recorded and made available to others. Collaboratively inspecting collections is crucial. Ontologies are an often preferred mechanism for modeling what is inside a collection, but this is probably limitative and narrow.

LINKMEDIA is concerned with making use of existing strategies in relation with ontologies and knowledge bases. In addition, LINKMEDIA uses mechanisms allowing to materialize the knowledge gradually acquired by humans and that might be subsequently used either by other humans or by computers in order to better and more precisely analyze collections. This line of work is instantiated at the core of the iCODA project LINKMEDIA coordinates. We are therefore concerned with:

- *Multimedia analysis and ontologies.* We develop approaches for linking multimedia content to entities in ontologies for text and images, building on results in multimodal embedding to cast entity linking into a nearest neighbor search problem in a high-dimensional joint embedding of content and entities [85]. We also investigate the use of ontological knowledge to facilitate information extraction from content [9].
- *Explainability and accountability in information extraction.* In relation with ontologies and entity linking, we develop innovative approaches to explain statistical relations found in data, in particular lexical or entity co-occurrences in textual data, for example using embeddings constrained with translation properties of RDF knowledge or path-based explanation within RDF graphs. We also work on confidence measures in entity linking and information extraction, studying how the notions of confidence and information source can be accounted for in knowledge basis and used in human-centric collaborative exploration of collections.
- *Dynamic evolution of models for information extraction.* In interactive exploration and information extraction, e.g., on cultural or educational material, knowledge progressively evolves as the process goes on, requiring on-the-fly design of new models for content-based information extractors from very few examples, as well as continuous adaptation of the models. Combining in a seamless way low-shot, active and incremental learning techniques is a key issue that we investigate to enable this dynamic mechanisms on selected applications.

3.2.1.4. Research Direction 2: Accessing Information

LINKMEDIA centers its activities on enabling humans to make good use of vast multimedia collections. This material takes all its cultural and economic value, all its artistic wonder when it can be accessed, watched, searched, browsed, visualized, summarized, classified, shared, ... This allows users to fully enjoy the incalculable richness of the collections. It also makes it possible for companies to create business rooted in this multimedia material.

Accessing the multimedia data that is inside a collection is complicated by the various type of data, their volume, their length, etc. But it is even more complicated to access the information that is not materialized in documents, such as the relationships between parts of different documents that however share some similarity. LINKMEDIA in its first four years of existence established itself as one of the leading teams in the field of multimedia analytics, contributing to the establishment of a dedicated community (refer to the various special sessions we organized with MMM, the iCODA and the LIMAH projects, as well as [66], [67], [63]).

Overall, facilitating the access to the multimedia material, to the relevant information and the corresponding knowledge asks for algorithms that efficiently *search* collections in order to identify the elements of collections or of the acquired knowledge that are matching a query, or that efficiently allow *navigating* the collections or the acquired knowledge. Navigation is likely facilitated if techniques are able to handle information and knowledge according to hierarchical perspectives, that is, allow to reveal data according to various levels of details. Aggregating or *summarizing* multimedia elements is not trivial.

Three topics are therefore in relation with this second research direction. LINKMEDIA tackles the issues in relation to searching, to navigating and to summarizing multimedia information. Information needs when discovering the content of a multimedia collection can be conveniently mapped to the exploration-search axis, as first proposed by Zahálka and Worring in [88], and illustrated by Figure 1 where expert users typically work near the right end because their tasks involve precise queries probing search engines. In contrast, lay-users start near the exploration end of the axis. Overall, users may alternate searches and explorations by going back and forth along the axis. The underlying model and system must therefore be highly dynamic, support interactions with the users and propose means for easy refinements. LINKMEDIA contributes to advancing the state of the art in searching operations, in navigating operations (also referred to as browsing), and in summarizing operations.

3.2.1.4.1. Searching.

Search engines must run similarity searches very efficiently. High-dimensional indexing techniques therefore play a central role. Yet, recent contributions in ML suggest to revisit indexing in order to adapt to the specific properties of modern features describing contents.

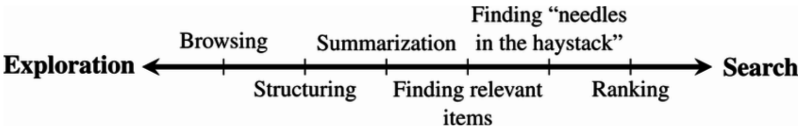


Figure 1. Exploration-search axis with example tasks

- *Advanced scalable indexing.* High-dimensional indexing is one of the foundations of LINKMEDIA. Modern features extracted from the multimedia material with the most recent ML techniques shall be indexed as well. This, however, poses a series of difficulties due to the dimensionality of these features, their possible sparsity, the complex metrics in use, the task in which they are involved (instance search, k-nn, class prototype identification, manifold search [65], time series retrieval, ...). Furthermore, truly large datasets require involving sketching [50], secondary storage and/or distribution [49], [48], alleviating the explosion of the number of features to consider due to their local nature or other innovative methods [64], all introducing complexities. Last, indexing multimodal embedded spaces poses a new series of challenges.
- *Improving quality.* Scalable indexing techniques are approximate, and what they return typically includes a fair amount of false positives. LINKMEDIA works on improving the quality of the results returned by indexing techniques. Approaches taking into account neighborhoods [59], manifold structures instead of pure distance based similarities [65] must be extended to cope with advanced indexing in order to enhance quality. This includes feature selection based on intrinsic dimensionality estimation [47].
- *Dynamic indexing.* Feature collections grow, and it is not an option to fully reindex from scratch an updated collection. This trivially applies to the features directly extracted from the media items, but also to the base class prototypes that can evolve due to the non-static nature of learning processes. LINKMEDIA will continue investigating what is at stake when designing dynamic indexing strategies.

3.2.1.4.2. Navigating.

Navigating a multimedia collection is very central to its understanding. It differs from searching as navigation is not driven by any specific query. Rather, it is mostly driven by the relationships that various documents have one another. Relationships are supported by the links between documents and/or parts of documents. Links rely on semantic similarity, depicting the fact that two documents share information on the same topic. But other aspects than semantics are also at stake, e.g., time with the dates of creation of the documents or geography with mentions or appearance in documents of some geographical landmarks or with geo-tagged data.

In multimedia collections, links can be either implicit or explicit, the latter being much easier to use for navigation. An example of an implicit link can be the name of someone existing in several different news articles; we, as humans, create a mental link between them. In some cases, the computer misses such configurations, leaving such links implicit. Implicit links are subject to human interpretation, hence they are sometimes hard to identify for any automatic analysis process. Implicit links not being materialized, they can therefore hardly be used for navigation or faceted search. Explicit links can typically be seen as hyperlinks, established either by content providers or, more aligned with LINKMEDIA, automatically determined from content analysis. Entity linking (linking content to an entity referenced in a knowledge base) is a good example of the creation of explicit links. Semantic similarity links, as investigated in the LIMAH project and as considered in the search and hyperlinking task at MediaEval and TRECVid, are also prototypical links that can be made explicit for navigation. Pursuing work, we investigate two main issues:

- *Improving multimodal content-based linking.* We exploit achievements in entity linking to go beyond lexical or lexico-visual similarity and to provide semantic links that are easy to interpret for humans; carrying on, we work on link characterization, in search of mechanisms addressing link explainability (i.e., what is the nature of the link), for instance using attention models so as to focus on the common parts of two documents or using natural language generation; a final topic that we address is that of linking textual content to external data sources in the field of journalism, e.g., leveraging topic models and cue phrases along with a short description of the external sources.
- *Dynamicity and user-adaptation.* One difficulty for explicit link creation is that links are often suited for one particular usage but not for another, thus requiring creating new links for each intended use; whereas link creation cannot be done online because of its computational cost, the alternative is to generate (almost) all possible links and provide users with selection mechanisms enabling personalization and user-adaptation in the exploration process; we design such strategies and investigate their impact on exploration tasks in search of a good trade-off between performance (few high-quality links) and genericity.

3.2.1.4.3. Summarizing.

Multimedia collections contain far too much information to allow any easy comprehension. It is mandatory to have facilities to aggregate and summarize a large body on information into a compact, concise and meaningful representation facilitating getting insight. Current technology suggests that multimedia content aggregation and story-telling are two complementary ways to provide users with such higher-level views. Yet, very few studies already investigated these issues. Recently, video or image captioning [87], [82] have been seen as a way to summarize visual content, opening the door to state-of-the-art multi-document text summarization [61] with text as a pivot modality. Automatic story-telling has been addressed for highly specific types of content, namely TV series [54] and news [73], [80], but still need a leap forward to be mostly automated, e.g., using constraint-based approaches for summarization [51], [80].

Furthermore, not only the original multimedia material has to be summarized, but the knowledge acquired from its analysis is also to summarize. It is important to be able to produce high-level views of the relationships between documents, emphasizing some structural distinguishing qualities. Graphs establishing such relationships need to be constructed at various level of granularity, providing some support for summarizing structural traits.

Summarizing multimedia information poses several scientific challenges that are:

- *Choosing the most relevant multimedia aggregation type*: Taking a multimedia collection into account, a same piece of information can be present in several modalities. The issue of selecting the most suitable one to express a given concept has thus to be considered together with the way to mix the various modalities into an acceptable production. Standard summarization algorithms have to be revisited so that they can handle continuous representation spaces, allowing them to benefit from the various modalities [55].

- *Expressing user's preferences*: Different users may appreciate quite different forms of multimedia summaries, and convenient ways to express their preferences have to be proposed. We for example focus on the opportunities offered by the constraint-based framework.
- *Evaluating multimedia summaries*: Finding criteria to characterize what a good summary is remains challenging, e.g., how to measure the global relevance of a multimodal summary and how to compare information between and across two modalities. We tackle this issue particularly via a collaboration with A. Smeaton at DCU, comparing the automatic measures we will develop to human judgments obtained by crowd-sourcing;
- *Taking into account structuring and dynamicity*: Typed links between multimedia fragments, and hierarchical topical structures of documents obtained via work previously developed within the team are two types of knowledge which have seldom been considered as long as summarization is concerned. Knowing that the event present in a document is causally related to another event described in another document can however modify the ways summarization algorithms have to consider information. Moreover the question of producing coarse-to-fine grain summaries exploiting the topical structure of documents is still an open issue. Summarizing dynamic collections is also challenging and it is one of the questions we consider.

4. Application Domains

4.1. Asset management in the entertainment business

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 bibliothèque Nationale de France, 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. This also concerns TV on-demand and replay services as well as social TV services and multi-screen applications. Enriching multimedia content, with explicit links targeting either multimedia material or knowledge databases is central here.

4.3. Data journalism

Data journalism forms an application domain where most of the technology developed by LINKMEDIA can be used. On the one hand, data journalists often need to inspect multiple heterogeneous information sources, some being well structured, some other being fully unstructured. They need to access (possibly their own) archives with either searching or navigational means. To gradually construct insight, they need collaborative multimedia analytics processes as well as elements of trust in the information they use as foundations for their investigations. Trust in the information, watching for adversarial and/or (deep) fake material, accountability are all crucial here.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Highlights of the year

- Our activities in relation with fake news were extensively highlighted in 2019. Ewa Kijak and Vincent Claveau gave a few interviews in newspapers, in a nationwide radio broadcast as well as in several TV shows.
- A chaire position in Artificial Intelligence for Defense has been granted to Teddy Furon. This chaire is supported by the national Defense Innovation Agency. The chaire will last 4 years, starting early 2020.
- Laurent Amsaleg (General Chair), Guillaume Gravier (Program Committee Chair), Yannis Avrithis (Workshops Chair) as well as almost all students of LINKMEDIA (as volunteers) were involved in running the 27th ACM Multimedia conference in Nice. This edition, very successful, was attended by close to 800 people.

5.1.2. Awards

Oriane Siméoni received the best presentation award from the International Computer Vision Summer School (ICVSS) 2019 [0].

6. New Software and Platforms

6.1. Lookinlabs-Global

KEYWORD: Search Engine

FUNCTIONAL DESCRIPTION: Lookinlabs allows you to find, among teams/individuals/publications, those best matching your query.

- Authors: William Kokou Dedzoe and Jean Hany
- Contact: Vincent Claveau

6.2. TagEx

Yet another Part-of-Speech Tagger for French

KEYWORD: Natural language processing

FUNCTIONAL DESCRIPTION: TagEx is available as a web-service on https://allgo.inria.fr . Refer to Allgo for its usage.

- Contact: Vincent Claveau
- URL: https://allgo.inria.fr/app/tagex

6.3. NegDetect

Negation Detection

KEYWORD: Natural language processing

FUNCTIONAL DESCRIPTION: NegDetect relies on several layers of machine learning techniques (CRF, neural networks).

- Contact: Vincent Claveau

[0] https://iplab.dmi.unict.it/icvss2019/

7. New Results

7.1. Extracting and Representing Information

7.1.1. Text Mining in the Clinical Domain

Participants: Clément Dalloux, Vincent Claveau.

Clinical records cannot be shared, which is a real hurdle to develop and compare information extraction techniques. In the framework of the BigClin Project we have developed annotated corpora, that share the same linguistic properties than records, but can be freely distributed for research purposes. Several corpora and several types of annotation were proposed for French, Portuguese and English. They are made freely available for research purposes and are described in [27], [25]. These corpora will foster reproducible research on clinical text mining.

Thanks to these datasets, we have organized the DeFT text-mining competition in 2019. Several NLP techniques and tools have been developed within the project in order to identify relevant medical or linguistic information [30], [26]. They are all chiefly based on machine learning approaches, and for most of them, more specifically, on deep learning. For instance, we have developed a new Part-of-Speech tagger and lemmatizer for French, especially suited to handle medical texts; it is freely available as a web-service at https://allgo.inria.fr . The identification of negation and uncertainty is important to precisely understand the clinical texts. Thus, we have continued our work on neural techniques to find the negation/uncertainty cues and their scope (part of sentence concerned by the negation or uncertainty). It achieves state-of-the-art results on English, and is pioneer work for French and Portuguese for which it sets a new standard [4], [21]; it is available at https://allgo.inria.fr . Other achievements in text-mining include: numerical value extraction (finding concepts that are measured, such as lab results, numerical expressions, their units) in French, English and Portuguese, the identification of gender, age, outcome and admission reasons in French clinical texts, ...

7.1.2. Embedding in hyperbolic spaces

Participants: François Torregrossa, Vincent Claveau, Guillaume Gravier.

During this year, we have studied non-Euclidean spaces into which one can embed data (for instance, words). We have developed the HierarX tool which projects multiple datasources into hyperbolic manifolds: Lorentz or Poincaré. From similarities between word pairs or continuous word representations in high dimensional spaces, HierarX is able to embed knowledge in hyperbolic geometries with small dimensionality. Those shape information into continuous hierarchies. The source code is available on the Inria's GitLab.

7.1.3. Aggregation and embedding for group membership verification

Participants: Marzieh Gheisari Khorasgani, Teddy Furon, Laurent Amsaleg.

This paper proposes a group membership verification protocol preventing the curious but honest server from reconstructing the enrolled signatures and inferring the identity of querying clients [24]. The protocol quantizes the signatures into discrete embeddings, making reconstruction difficult. It also aggregates multiple embeddings into representative values, impeding identification. Theoretical and experimental results show the trade-off between the security and error rates.

7.1.4. Group Membership Verification with Privacy: Sparse or Dense?

Participants: Marzieh Gheisari Khorasgani, Teddy Furon, Laurent Amsaleg.

Group membership verification checks if a biometric trait corresponds to one member of a group without revealing the identity of that member. Recent contributions provide privacy for group membership protocols through the joint use of two mechanisms: quantizing templates into discrete embeddings, and aggregating several templates into one group representation. However, this scheme has one drawback: the data structure representing the group has a limited size and cannot recognize noisy query when many templates are aggregated. Moreover, the sparsity of the embeddings seemingly plays a crucial role on the performance verification. This contribution proposes a mathematical model for group membership verification allowing to reveal the impact of sparsity on both security, compactness, and verification performances [23]. This models bridges the gap towards a Bloom filter robust to noisy queries. It shows that a dense solution is more competitive unless the queries are almost noiseless.

7.1.5. Privacy Preserving Group Membership Verification and Identification

Participants: Marzieh Gheisari Khorasgani, Teddy Furon, Laurent Amsaleg.

When convoking privacy, group membership verification checks if a biometric trait corresponds to one member of a group without revealing the identity of that member. Similarly, group membership identification states which group the individual belongs to, without knowing his/her identity. A recent contribution provides privacy and security for group membership protocols through the joint use of two mechanisms: quantizing biometric templates into discrete embeddings, and aggregating several templates into one group representation. This paper significantly improves that contribution because it jointly learns how to embed and aggregate instead of imposing fixed and hard coded rules [10]. This is demonstrated by exposing the mathematical underpinnings of the learning stage before showing the improvements through an extensive series of experiments targeting face recognition. Overall, experiments show that learning yields an excellent trade-off between security/privacy and the verification/identification performances.

7.1.6. Intrinsic Dimensionality Estimation within Tight Localities

Participants: Laurent Amsaleg, Oussama Chelly [Microsoft Germany], Michael Houle [National Institute of Informatics, Japan], Ken-Ichi Kawarabayashi [National Institute of Informatics, Japan], Miloš Radovanović [Univ. Novi Sad, Serbia], Weeris Treeratanajaru [Chulalongkorn University, Thailand].

Accurate estimation of Intrinsic Dimensionality (ID) is of crucial importance in many data mining and machine learning tasks, including dimensionality reduction, outlier detection, similarity search and subspace clustering. However, since their convergence generally requires sample sizes (that is, neighborhood sizes) on the order of hundreds of points, existing ID estimation methods may have only limited usefulness for applications in which the data consists of many natural groups of small size. In this paper, we propose a local ID estimation strategy stable even for 'tight' localities consisting of as few as 20 sample points [31]. The estimator applies MLE techniques over all available pairwise distances among the members of the sample, based on a recent extreme-value-theoretic model of intrinsic dimensionality, the Local Intrinsic Dimension (LID). Our experimental results show that our proposed estimation technique can achieve notably smaller variance, while maintaining comparable levels of bias, at much smaller sample sizes than state-of-the-art estimators.

7.1.7. Selective Biogeography-Based Optimizer Considering Resource Allocation for Large-Scale Global Optimization

Participants: Meiji Cui [Tongji University, China], Li Li [Tongji University, China], Miaojing Shi.

Biogeography-based optimization (BBO), a recent proposed meta-heuristic algorithm, has been successfully applied to many optimization problems due to its simplicity and efficiency. However, BBO is sensitive to the curse of dimensionality; its performance degrades rapidly as the dimensionality of the search space increases. In [3], a selective migration operator is proposed to scale up the performance of BBO and we name it selective BBO (SBBO). The differential migration operator is selected heuristically to explore the global area as far as possible whist the normal distributed migration operator is chosen to exploit the local area. By the means of heuristic selection, an appropriate migration operator can be used to search the global optimum efficiently. Moreover, the strategy of cooperative co-evolution (CC) is adopted to solve large-scale global optimization problems (LSOPs). To deal with subgroup imbalance contribution to the whole solution in the context of CC, a more efficient computing resource allocation is proposed. Extensive experiments are conducted on the CEC 2010 benchmark suite for large-scale global optimization, and the results show the effectiveness and efficiency of SBBO compared with BBO variants and other representative algorithms for LSOPs. Also, the results confirm that the proposed computing resource allocation is vital to the large-scale optimization within the limited computation budget.

7.1.8. Friend recommendation for cross marketing in online brand community based on intelligent attention allocation link prediction algorithm

Participants: Shugang Li [Shanghai University, China], Xuewei Song [Shanghai University, China], Hanyu Lu [Shanghai University, China], Linyi Zeng [Shanghai University, Industrial and Commercial Bank of China, China], Miaojing Shi, Fang Liu [Shanghai University, China].

Circle structure of online brand communities allows companies to conduct cross-marketing activities by the influence of friends in different circles and build strong and lasting relationships with customers. However, existing works on the friend recommendation in social network do not consider establishing friendships between users in different circles, which has the problems of network sparsity, neither do they study the adaptive generation of appropriate link prediction algorithms for different circle features. In order to fill the gaps in previous works, the intelligent attention allocation link prediction algorithm is proposed to adaptively build attention allocation index (AAI) according to the sparseness of the network and predict the possible friendships between users in different circles. The AAI reflects the amount of attention allocated to the user pair by their common friend in the triadic closure structure, which is decided by the friend count of the common friend. Specifically, for the purpose of overcoming the problem of network sparsity, the AAIs of both the direct common friends and indirect ones are developed. Next, the decision tree (DT) method is constructed to adaptively select the suitable AAIs for the circle structure based on the density of common friends and the dispersion level of common friends' attention. In addition, for the sake of further improving the accuracy of the selected AAI, its complementary AAIs are identified with support vector machine model according to their similarity in value, direction, and ranking. Finally, the mutually complementary indices are combined into a composite one to comprehensively portray the attention distribution of common friends of users in different circles and predict their possible friendships for cross-marketing activities. Experimental results on Twitter and Google+ show that the model has highly reliable prediction performance [5].

7.1.9. Revisiting the medial axis for planar shape decomposition

Participants: Nikos Papanelopoulos [NTUA, Greece], Yannis Avrithis, Stefanos Kollias [U. of Lincoln, UK].

We present a simple computational model for planar shape decomposition that naturally captures most of the rules and salience measures suggested by psychophysical studies, including the minima and short-cut rules, convexity, and symmetry. It is based on a medial axis representation in ways that have not been explored before and sheds more light into the connection between existing rules like minima and convexity. In particular, vertices of the exterior medial axis directly provide the position and extent of negative minima of curvature, while a traversal of the interior medial axis directly provides a small set of candidate endpoints for part-cuts. The final selection follows a prioritized processing of candidate part-cuts according to a local convexity rule that can incorporate arbitrary salience measures. Neither global optimization nor differentiation is involved. We provide qualitative and quantitative evaluation and comparisons on ground-truth data from psycho-physical experiments. With our single computational model, we outperform even an ensemble method on several other competing models [6].

7.1.10. Graph-based Particular Object Discovery

Participants: Oriane Siméoni, Ahmet Iscen [Univ. Prague], Giorgos Tolias [Univ. Prague], Yannis Avrithis, Ondra Chum [Univ. Prague].

Severe background clutter is challenging in many computer vision tasks, including large-scale image retrieval. Global descriptors, that are popular due to their memory and search efficiency, are especially prone to corruption by such a clutter. Eliminating the impact of the clutter on the image descriptor increases the chance of retrieving relevant images and prevents topic drift due to actually retrieving the clutter in the case of query expansion. In this work, we propose a novel salient region detection method. It captures, in an unsupervised manner, patterns that are both discriminative and common in the dataset. Saliency is based on a centrality measure of a nearest neighbor graph constructed from regional CNN representations of dataset images. The proposed method exploits recent CNN architectures trained for object retrieval to construct the image representation from the salient regions. We improve particular object retrieval on challenging datasets containing small objects [7].

7.1.11. Label Propagation for Deep Semi-supervised Learning

Participants: Ahmet Iscen [Univ. Prague], Giorgos Tolias [Univ. Prague], Yannis Avrithis, Ondra Chum [Univ. Prague].

Semi-supervised learning is becoming increasingly important because it can combine data carefully labeled by humans with abundant unlabeled data to train deep neural networks. Classic methods on semi-supervised learning that have focused on transductive learning have not been fully exploited in the inductive framework followed by modern deep learning. The same holds for the manifold assumption—that similar examples should get the same prediction. In this work, we employ a transductive label propagation method that is based on the manifold assumption to make predictions on the entire dataset and use these predictions to generate pseudo-labels for the unlabeled data and train a deep neural network. At the core of the transductive method lies a nearest neighbor graph of the dataset that we create based on the embeddings of the same network. Therefore our learning process iterates between these two steps. We improve performance on several datasets especially in the few labels regime and show that our work is complementary to current state of the art [12], [38].

7.1.12. Dense Classification and Implanting for Few-Shot Learning

Participants: Yann Lifchitz, Yannis Avrithis, Sylvaine Picard [SAFRAN Group], Andrei Bursuc [Valéo].

Few-shot learning for deep neural networks is a highly challenging and key problem in many computer vision tasks. In this context, we are targeting knowledge transfer from a set with abundant data to other sets with few available examples. We propose in [14], [40] two simple and effective solutions: (i) dense classification over feature maps, which for the first time studies local activations in the domain of few- shot learning, and (ii) implanting, that is, attaching new neurons to a previously trained network to learn new, task-specific features. Implanting enables training of multiple layers in the few-shot regime, departing from most related methods derived from metric learning that train only the final layer. Both contributions show consistent gains when used individually or jointly and we report state of the art performance on few-shot classification on miniImageNet.

7.1.13. Point in, Box out: Beyond Counting Persons in Crowds

Participants: Yuting Liu [Sichuan University, China], Miaojing Shi, Qijun Zhao [Sichuan University, China], Xiaofang Wang [RAINBOW Team, IRISA].

Modern crowd counting methods usually employ deep neural networks (DNN) to estimate crowd counts via density regression. Despite their significant improvements, the regression-based methods are incapable of providing the detection of individuals in crowds. The detection-based methods, on the other hand, have not been largely explored in recent trends of crowd counting due to the needs for ex- pensive bounding box annotations. In this work, we instead propose a new deep detection network with only point supervision required [15]. It can simultaneously detect the size and location of human heads and count them in crowds. We first mine useful person size information from point-level annotations and initialize the pseudo ground truth bounding boxes. An online updating scheme is introduced to refine the pseudo ground truth during training; while a locally- constrained regression loss is designed to provide additional constraints on the size of the predicted boxes in a local neighborhood. In the end, we propose a curriculum learning strategy to train the network from images of relatively accurate and easy pseudo ground truth first. Extensive experiments are conducted in both detection and counting tasks on several standard benchmarks, e.g. ShanghaiTech, UCF CC 50, WiderFace, and TRANCOS datasets, and the results show the superiority of our method over the state-of-the-art.

7.1.14. Revisiting Perspective Information for Efficient Crowd Counting

Participants: Miaojing Shi, Zhaohui Yang [Peking University, China], Chao Xu [Peking University, China], Qijun Chen [Tongji University, China].

Crowd counting is the task of estimating people numbers in crowd images. Modern crowd counting methods employ deep neural networks to estimate crowd counts via crowd density regressions. A major challenge of this task lies in the perspective distortion, which results in drastic person scale change in an image.

Density regression on the small person area is in general very hard. In this work, we propose a perspective-aware convolutional neural network (PACNN) for efficient crowd counting, which integrates the perspective information into density regression to provide additional knowledge of the person scale change in an image [18]. Ground truth perspective maps are firstly generated for training; PACNN is then specifically designed to predict multi-scale perspective maps, and encode them as perspective-aware weighting layers in the network to adaptively combine the outputs of multi-scale density maps. The weights are learned at every pixel of the maps such that the final density combination is robust to the perspective distortion. We conduct extensive experiments on the ShanghaiTech, WorldExpo'10, UCF CC 50, and UCSD datasets, and demonstrate the effectiveness and efficiency of PACNN over the state-of-the-art.

7.1.15. Local Features and Visual Words Emerge in Activations

Participants: Oriane Siméoni, Yannis Avrithis, Ondra Chum [Univ. Prague].

We propose a novel method of deep spatial matching (DSM) for image retrieval [19], [41]. Initial ranking is based on image descriptors extracted from convolutional neural network activations by global pooling, as in recent state-of- the-art work. However, the same sparse 3D activation tensor is also approximated by a collection of local features. These local features are then robustly matched to approximate the optimal alignment of the tensors. This hap- pens without any network modification, additional layers or training. No local feature detection happens on the original image. No local feature descriptors and no visual vocabulary are needed throughout the whole process. We experimentally show that the proposed method achieves the state-of-the-art performance on standard benchmarks across different network architectures and different global pooling methods. The highest gain in performance is achieved when diffusion on the nearest-neighbor graph of global descriptors is initiated from spatially verified images.

7.1.16. Combining convolutional side-outputs for road image segmentation

Participants: Raquel Almeida, Simon Malinowski, Ewa Kijak, Silvio Guimaraes [PUC Minas].

Image segmentation consists in creating partitions within an image into meaningful areas and objects. It can be used in scene understanding and recognition, in fields like biology, medicine, robotics, satellite imaging, amongst others. In this work [17], we take advantage of the learned model in a deep architecture, by extracting side-outputs at different layers of the network for the task of image segmentation. We study the impact of the amount of side-outputs and evaluate strategies to combine them. A post-processing filtering based on mathematical morphology idempotent functions is also used in order to remove some undesirable noises. Experiments were performed on the publicly available KITTI Road Dataset for image segmentation. Our comparison shows that the use of multiples side outputs can increase the overall performance of the network, making it easier to train and more stable when compared with a single output in the end of the network. Also, for a small number of training epochs (500), we achieved a competitive performance when compared to the best algorithm in KITTI Evaluation Server.

7.1.17. BRIEF-based mid-level representations for time series classification

Participants: Raquel Almeida, Simon Malinowski, Silvio Guimaraes [PUC Minas].

Time series classification has been widely explored over the last years. Amongst the best approaches for that task, many are based on the Bag-of-Words framework, in which time series are transformed into a histogram of word occurrences. These words represent quantized features that are extracted beforehand. In this work [20], we aim to evaluate the use of accurate mid-level representation called BossaNova in order to enhance the Bag-of-Words representation and to propose a new binary time series descriptor, called BRIEF-based descriptor. More precisely, this kind of representation enables to reduce the loss induced by feature quantization. Experiments show that this representation in conjunction to BRIEF-based descriptor is statistically equivalent to traditional Bag-of-Words, in terms time series classification accuracy, being about 4 times faster. Furthermore, it is very competitive when compared to the state-of-the-art.

7.1.18. Toward a Framework for Seasonal Time Series Forecasting Using Clustering

Participants: Simon Malinowski, Thomas Guyet [LACODAM Team], Colin Leverger [LACODAM Team], Alexandre Termier [LACODAM Team].

Seasonal behaviours are widely encountered in various applications. For instance, requests on web servers are highly influenced by our daily activities. Seasonal forecasting consists in forecasting the whole next season for a given seasonal time series. It may help a service provider to provision correctly the potentially required resources, avoiding critical situations of over- or under provision. In this article, we propose a generic framework to make seasonal time series forecasting. The framework combines machine learning techniques (1) to identify the typical seasons and (2) to forecast the likelihood of having a season type in one season ahead. We study in[13] this framework by comparing the mean squared errors of forecasts for various settings and various datasets. The best setting is then compared to state-of-the-art time series forecasting methods. We show that it is competitive with them.

7.1.19. Smooth Adversarial Examples

Participants: Hanwei Zhang, Yannis Avrithis, Teddy Furon, Laurent Amsaleg.

This paper investigates the visual quality of the adversarial examples. Recent papers propose to smooth the perturbations to get rid of high frequency artefacts. In this work, smoothing has a different meaning as it perceptually shapes the perturbation according to the visual content of the image to be attacked [44]. The perturbation becomes locally smooth on the flat areas of the input image, but it may be noisy on its textured areas and sharp across its edges. This operation relies on Laplacian smoothing, well-known in graph signal processing, which we integrate in the attack pipeline. We benchmark several attacks with and without smoothing under a white-box scenario and evaluate their transferability. Despite the additional constraint of smoothness, our attack has the same probability of success at lower distortion.

7.1.20. Walking on the Edge: Fast, Low-Distortion Adversarial Examples

Participants: Hanwei Zhang, Yannis Avrithis, Teddy Furon, Laurent Amsaleg.

Adversarial examples of deep neural networks are receiving ever increasing attention because they help in understanding and reducing the sensitivity to their input. This is natural given the increasing applications of deep neural networks in our everyday lives. When white-box attacks are almost always successful, it is typically only the distortion of the perturbations that matters in their evaluation. In this work [45], we argue that speed is important as well, especially when considering that fast attacks are required by adversarial training. Given more time, iterative methods can always find better solutions. We investigate this speed-distortion trade-off in some depth and introduce a new attack called boundary projection (BP) that improves upon existing methods by a large margin. Our key idea is that the classification boundary is a manifold in the image space: we therefore quickly reach the boundary and then optimize distortion on this manifold.

7.1.21. Accessing watermarking information: Error exponents in the noisy case

Participant: Teddy Furon.

The study of the error exponents of zero-bit watermarking is addressed in the article by Comesana, Merhav, and Barni, under the assumption that the detector relies solely on second order joint empirical statistics of the received signal and the watermark. This restriction leads to the well-known dual hypercone detector, whose score function is the absolute value of the normalized correlation. They derive the false negative error exponent and the optimum embedding rule. However, they only focus on high SNR regime, i.e. the noiseless scenario. This work extends this theoretical study to the noisy scenario. It introduces a new definition of watermarking robustness based on the false negative error exponent, derives this quantity for the dual hypercone detector, and shows that its performances is almost equal to Costa's lower bound [22].

7.1.22. Detecting fake news and image forgeries

Participants: Cédric Maigrot, Vincent Claveau, Ewa Kijak.

Social networks make it possible to share information rapidly and massively. Yet, one of their major drawback comes from the absence of verification of the piece of information, especially with viral messages. Based on the work already presented in the previous years, C. Maigrot defended his thesis on the detection of image forgeries, classification of reinformation websites, and on the late fusion of models based on the text, image and source analysis [1]. This work was also given a large visibility thanks to numerous interviews in Press and TV (see the dedicated section about popularization).

7.1.23. Learning Interpretable Shapelets for Time Series Classification through Adversarial Regularization

Times series classification can be successfully tackled by jointly learning a shapelet-based representation of the series in the dataset and classifying the series according to this representation. However, although the learned shapelets are discriminative, they are not always similar to pieces of a real series in the dataset. This makes it difficult to interpret the decision, i.e. difficult to analyze if there are particular behaviors in a series that triggered the decision. In this work [29], we make use of a simple convolutional network to tackle the time series classification task and we introduce an adversarial regularization to constrain the model to learn more interpretable shapelets. Our classification results on all the usual time series benchmarks are comparable with the results obtained by similar state-of-the-art algorithms but our adversarially regularized method learns shapelets that are, by design, interpretable.

7.1.24. Using Knowledge Base Semantics in Context-Aware Entity Linking

Participants: Cheikh Brahim El Vaigh, Guillaume Gravier, Pascale Sébillot.

Done as part of the IPL iCODA, in collaboration with CEDAR Inria team.

Entity linking is a core task in textual document processing, which consists in identifying the entities of a knowledge base (KB) that are mentioned in a text. Approaches in the literature consider either independent linking of individual mentions or collective linking of all mentions. Regardless of this distinction, most approaches rely on the Wikipedia encyclopedic KB in order to improve the linking quality, by exploiting its entity descriptions (web pages) or its entity interconnections (hyperlink graph of web pages). We devised a novel collective linking technique which departs from most approaches in the literature by relying on a structured RDF KB [9]. This allows exploiting the semantics of the interrelationships that candidate entities may have at disambiguation time rather than relying on raw structural approximation based on Wikipedia's hyperlink, graph. The few approaches that also use an RDF KB simply rely on the existence of a relation between the candidate entities to which mentions may be linked. Instead, we weight such relations based on the RDF KB structure and propose an efficient decoding strategy for collective linking. Experiments on standard benchmarks show significant improvement over the state of the art.

7.1.25. Neural-based lexico-syntactic relation extraction in news archives

Participants: Guillaume Gravier, Cyrielle Mallart, Pascale Sébillot.

Done as part of the IPL iCODA, in collaboration with Ouest France

Relation extraction is the task of finding and classifying the relationship between two entities in a text. We pursued work on the detection of relations between entities, seen as a binary classification problem. In the context of large-scale news archives, we argue that detection is paramount before even considering classification, where most approaches consider the two tasks jointly with a null garbage class. This does hardly allow for the detection of relations for unseen categories, which are all considered as garbage. We designed a bi-LSTM sequence neural model acting on features extracted from the surface realization, the part-of-speech tags and the dependency parse tree and compared with a state-of-the-art relation detection LSTM-based approach. Experimental evaluations rely on a dataset derived from 200k Wikipedia articles in French containing 4M linked mentions of entities: 330k pairs of entities co-occur in the same sentence, of which 1 % are actual relations according to Wikidata. Results show the benefit of our binary detection approach over previous methods and over joint detection and classification.

7.1.26. Graph Convolutional Networks for Learning with Few Clean and Many Noisy Labels

Participants: Ahmet Iscen [Google Research], Giorgos Tolias [Univ. Prague], Yannis Avrithis, Ondra Chum [Univ. Prague], Cordelia Schmid [Google Research].

In this work we consider the problem of learning a classifier from noisy labels when a few clean labeled examples are given [39]. The structure of clean and noisy data is modeled by a graph per class and Graph Convolutional Networks (GCN) are used to predict class relevance of noisy examples. For each class, the GCN is treated as a binary classifier learning to discriminate clean from noisy examples using a weighted binary cross-entropy loss function, and then the GCN-inferred "clean" probability is exploited as a relevance measure. Each noisy example is weighted by its relevance when learning a classifier for the end task. We evaluate our method on an extended version of a few-shot learning problem, where the few clean examples of novel classes are supplemented with additional noisy data. Experimental results show that our GCN-based cleaning process significantly improves the classification accuracy over not cleaning the noisy data and standard few-shot classification where only few clean examples are used. The proposed GCN-based method outperforms the transductive approach (Douze et al., 2018) that is using the same additional data without labels.

7.1.27. Rethinking deep active learning: Using unlabeled data at model training

Participants: Oriane Siméoni, Mateusz Budnik, Yannis Avrithis, Guillaume Gravier.

Active learning typically focuses on training a model on few labeled examples alone, while unlabeled ones are only used for acquisition. In this work we depart from this setting by using both labeled and unlabeled data during model training across active learning cycles [42]. We do so by using unsupervised feature learning at the beginning of the active learning pipeline and semi-supervised learning at every active learning cycle, on all available data. The former has not been investigated before in active learning, while the study of latter in the context of deep learning is scarce and recent findings are not conclusive with respect to its benefit. Our idea is orthogonal to acquisition strategies by using more data, much like ensemble methods use more models. By systematically evaluating on a number of popular acquisition strategies and datasets, we find that the use of unlabeled data during model training brings a spectacular accuracy improvement in image classification, compared to the differences between acquisition strategies. We thus explore smaller label budgets, even one label per class.

7.1.28. Training Object Detectors from Few Weakly-Labeled and Many Unlabeled Images

Participants: Zhaohui Yang [Peking University], Miaojing Shi, Yannis Avrithis, Chao Xu [Peking University], Vittorio Ferrari [Google Research].

Weakly-supervised object detection attempts to limit the amount of supervision by dispensing the need for bounding boxes, but still assumes image-level labels on the entire training set are available. In this work, we study the problem of training an object detector from one or few clean images with image-level labels and a larger set of completely unlabeled images [43]. This is an extreme case of semi-supervised learning where the labeled data are not enough to bootstrap the learning of a classifier or detector. Our solution is to use a standard weakly-supervised pipeline to train a student model from image-level pseudo-labels generated on the unlabeled set by a teacher model, bootstrapped by region-level similarities to clean labeled images. By using the recent pipeline of PCL and more unlabeled images, we achieve performance competitive or superior to many state of the art weakly-supervised detection solutions.

7.2. Accessing Information

7.2.1. Ontological modeling of human reading experience

Participants: Guillaume Gravier, Pascale Sébillot.

Done as part of the JPI CH READ-IT projects, in collaboration with Open University (UK) and Université Le Mans (FR)

Diaries, correspondence and authors' libraries provide important evidence into the evolution of ideas and society. Studying these phenomena is connected to understanding changes of perspective and values. Within the framework of the READ-IT project, we developed an ontological data approach modelling changes in the contents of diaries, correspondence and authors' libraries related to reading. By considering these three types of sources, we designed a conceptual data model to permit the study and increase the usability of sources containing evidence of reading experiences, highlighting common challenges and patterns related to changes to readers and to the medium of reading when confronting historical events [36], [8].

7.2.2. Integration of Exploration and Search: A Case Study of the M^3 Model

Participants: Snorri Gíslason [IT Univ. Copenhagen], Björn Þór Jónsson [IT Univ. Copenhagen], Laurent Amsaleg.

Effective support for multimedia analytics applications requires exploration and search to be integrated seamlessly into a single interaction model. Media metadata can be seen as defining a multidimensional media space, casting multimedia analytics tasks as exploration, manipulation and augmentation of that space. We present an initial case study of integrating exploration and search within this multidimensional media space [11]. We extend the M^3 model, initially proposed as a pure exploration tool, and show that it can be elegantly extended to allow searching within an exploration context and exploring within a search context. We then evaluate the suitability of relational database management systems, as representatives of today's data management technologies, for implementing the extended M^3 model. Based on our results, we finally propose some research directions for scalability of multimedia analytics.

7.2.3. Exquisitor: Breaking the Interaction Barrier for Exploration of 100 Million Images

Participants: Hanna Ragnarsdóttir [Reykjavik University], Þórhildur Þorleiksdóttir [Reykjavik University], Omar Shahbaz Khan [IT Univ. Copenhagen], Björn Þór Jónsson [IT Univ. Copenhagen], Gylfi Þór Gudmundsson [School of Computer Science, Reykjavik], Jan Zahálka [bohem.ai], Stevan Rudinac [University of Amsterdam], Laurent Amsaleg, Marcel Worring [University of Amsterdam].

We present Exquisitor, a media explorer capable of learning user preferences in real-time during interactions with the 99.2 million images of YFCC100M. Exquisitor owes its efficiency to innovations in data representation, compression, and indexing. Exquisitor can complete each interaction round, including learning preferences and presenting the most relevant results, in less than 30 ms using only a single CPU core and modest RAM. In short, Exquisitor can bring large-scale interactive learning to standard desktops and laptops, and even high-end mobile devices [16].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. CIFRE PhD: Incremental dynamic construction of knowledge bases from text mining

Participants: Guillaume Gravier, Cyrielle Mallart, Pascale Sébillot.

Duration: 3 years, started in Dec. 2018
Partner: Ouest France

In the context of a newspaper, the thesis explores the combination of text mining and knowledge representation techniques to assist the extraction, interpretation and validation of valuable pieces of information from the journal's content so as to incrementally build a full-scale knowledge base. This thesis is in close relation with the iCODA Inria Project Lab, with direct contribution to the project's results.

8.1.2. CIFRE PhD: Embedding heterogeneous data for directory search

Participants: Vincent Claveau, Guillaume Gravier, François Torregrossa.

Duration: 3 years, started in Dec. 2018

Partner: SoLocal

The thesis aims at learning how to jointly exploit heterogeneous sources of information (e.g., names, activity sector, user profiles, queries, etc.) in the design of neural network embeddings for information retrieval and language understanding. Applications cover natural language query analysis and personalized information retrieval in Pagesjaunes' directory.

8.1.3. CIFRE PhD: Few shot learning for object recognition in aerial images

Participants: Yannis Avrithis, Yann Lifchitz.

Duration: 3 years, started in March 2018
Partner: Safran Tech

This is a CIFRE PhD thesis project aiming to study architectures and learning techniques most suitable for object recognition from few samples and to validate these approaches on multiple recognition tasks and use-cases related to aerial images.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Computer vision for smart phones (MobilAI)

Participants: Yannis Avrithis, Mateusz Budnik.

Duration: 2 years, started in September 2018
Partners: Lamark, Quai des Apps, AriadNext

The ability of our mobile devices to process visual information is currently not limited by their camera or computing power but by the network. Many mobile apps suffer from long latency due to data transmitted over the network for visual search. MobilAI aims to provide fast visual recognition on mobile devices, offering quality user experience whatever the network conditions. The idea is to transfer efficient deep learning solutions for image classification and retrieval onto embedded platforms such as smart phones. The intention is to use such solutions in B2B and B2C application contexts, for instance recognizing products and ordering online, accessing information about artifacts in exhibitions, or identifying identity documents. In all cases, visual recognition is performed on the device, with minimal or no access to the network.

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: https://bigclin.cominlabs.u-bretagneloire.fr/fr

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. Inria Project Lab Knowledge-driven data and content collaborative analytics (iCODA)

Participants: Laurent Amsaleg, Cheikh Brahim El Vaigh, Guillaume Gravier, Cyrielle Mallart, 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.

9.2.2. Inria-BNF: Classification d'images patrimoniales (CIP)

Participants: Florent Michel, Laurent Amsaleg, Guillaume Gravier, Ewa Kijak, Yannis Avrithis.

Duration: 1 year, started in Dec 2018

This project is within the context of the collaborations between Inria and the French Ministry of Culture. In that context, we have started a collaboration with the French National Library (BNF) which collects, preserves and makes known the national documentary heritage. This collaboration aims at facilitating the automatic classification of heritage images through the use of recent deep-learning techniques. Such images are quite specific: they are not at all similar with what deep-learning techniques are used to work with, that is, the classification of heritage images does not target modern categories such as planes, cars, cats and dogs because this is irrelevant and because heritage collections do not include images of contemporary objects. Furthermore, heritage images come in vast quantities, but they are little annotated and deep-learning techniques can hardly rely on massive annotations to easily learn. Last, the learning has to be continuous as curators may need to add or modify existing classes, without re-learning everything from scratch.

The techniques of choice to reach that goal include the semi-supervised learning, low-shot learning techniques, knowledge transfer, fine tuning existing models, etc.

9.2.3. ANR Archival: Multimodal machine comprehension of language for new intelligent interfaces of scientific and cultural mediation

Participants: Laurent Amsaleg, Guillaume Gravier, Pascale Sébillot.

Duration: 3.5 year, started in Dec. 2019

The multidisciplinary and multi-actor ARCHIVAL project aims at yielding collaborations between researchers from the fields of Information and Communication Sciences as well as Computer Sciences around archive value enhancing and knowledge sharing for arts, culture and heritage. The project is structured around the following questionings: What part can machine comprehension methods play towards the reinterpretation of thematic archive collections? How can content mediation interfaces exploit results generated by current AI approaches?

ARCHIVAL teams will explore heterogeneous document collection structuration in order to explicitly reveal implicit links, to explain the nature of these links and to promote them in an intelligible way towards ergonomic mediation interfaces that will guarantee a successful appropriation of contents. A corpus has been delimited from the FMSH "self-management" collection, recently awarded as Collex, which will be completed from the large Canal-U academic audiovisual portal. The analysis and enhancement of this collection is of particular interest for Humanities and Social Sciences in a context where it becomes a necessity to structurally reconsider new models of socioeconomic development (democratic autonomy, social and solidarity-based economy, alternative development,. . .).

9.3. European Initiatives

9.3.1. EIT Digital CREEP2

Program: EIT Digital

Project acronym: CREEP 2

Project title: Cyberbullying effects prevention

Duration: 01/2019 - 12/2019

Coordinator: FBK, Italy

Other partners: Expert Systems (IT), Inria (FR), Engineering (IT)

Abstract: Project CREEP (Cyberbulling Effects Prevention) aims at identifying and preventing the possible negative impacts of cyberbullying on young people. It seeks to realise advanced technologies for the early detection of cyberbullying phenomena through the monitoring of social media and the communication of preventive advices and personalized recommendations tailored to adolescents' needs through a virtual coaching system (chatbot).

9.3.2. JPI CH READ-IT

Program: Joint Programming Initiative on Cultural Heritage

Project acronym: READ-IT

Project title: Reading Europe Advanced Data Investigation Tool

Duration: 05/2018 - 04/2021

Coordinator: Université Le Mans (FR)

Other partners: CNRS-IRISA (FR), Open University (UK), Universiteit Utrecht (NL), Institute of Czech Litterature (CZ)

Abstract: READ-IT is a transnational, interdisciplinary R&D project that will build a unique large-scale, user- friendly, open access, semantically-enriched investigation tool to identify and share groundbreaking evidence about 18th-21st century Cultural Heritage of reading in Europe. READ-IT will ensure the sustainable and reusable aggregation of qualitative data allowing an in-depth analysis of the Cultural Heritage of reading. State-of-the art technology in Semantic Web and information systems will provide a versatile, end-users oriented environment enabling scholars and ordinary readers to retrieve information from a vast amount of community-generated digital data leading to new understanding about the circumstances and effects of reading in Europe.

9.3.3. CHIST-ERA ID_IOT

Program: CHIST ERA

Project acronym: ID_IOT

Project title: Identification for the Internet of things

Duration: 3 years, started in Oct 2016.

Coordinator: Boris Skoric (Eindhoven Univ. of Technology (NL))

Other partners: Inria-RBA (Teddy Furon, Marzieh Gheisari Khorasgani), Univ. of Geneva (CH)

Abstract: 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 privacy preserving 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.

9.3.4. Collaborations with Major European Organizations

Program: ConFAP-CNRS Project

Project acronym: FIGTEM

Project title: FIne-Grain TExt Mining for clinical data

Duration: 01/2016 - 05/2019

Coordinator: CNRS-IRISA

Other partners: PUCPR, Curitiba, Brasil; CNRS-STL Lille; Inserm LTSI/CHU Rennes

Abstract: FIGTEM is a research project that involves STL-CNRS, CHU Rennes, PUC Parana, Curitiba and led by LINKMEDIAThis project aimed at developing natural language processing methods, including information extraction and indexing, dedicated to the clinical trial domain. The goal was to populate a formal representation of patients (via their electronic patient records) and clinical trial data in different languages (French, English, Portuguese). The main outcomes of the project was NLP tools for these 3 languages and annotated datasets made available for research purposes. It ended in May 2019.

9.4. International Initiatives

9.4.1. Inria International Partners

9.4.1.1. Informal International Partners

- Michael Houle, NII, Japan
- Marcel Worring, UvA, Netherlands
- Martha Larson, Radboud U., Netherlands

9.4.2. Participation in Other International Programs

9.4.2.1. STIC AmSud TRANSFORM

Program: STIC AmSud

Project acronym: TRANSFORM

Project title: Transforming multimedia data for indexing and retrieval purposes

Duration: 01/2018 - 31/2019

Partners: CNRS-IRISA (FR), PUC Minas (BR), UChile (CL)

9.4.2.2. CAPES/COFECUB HIMMD

Program: CAPES/COFECUB

Project acronym: HIMMD

Project title: Hierarchical graph-based analysis of image, video, and multimedia data

Duration: 01/2019 - 31/2022

Partners: LIGM (FR), IRISA (FR), INPG (FR), PUC Minas (BR), UNICamp (BR), UFMG (BR)

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- Silvio Guimaraes (PUC Minas, Brazil) visited the team in July (1 week). His visit was related to the Stic-Amsud project.
- Benjamin Bustos (Univ. Chile, Chile) visited the team in July (1 week). His visit was related to the Stic-Amsud project.

9.5.2. Visits to International Teams

9.5.2.1. Research Stays Abroad

- Yannis Avrithis, National and Kapodistrian University of Athens, 3 visits on April (1 week), June (1 week) and September-October 2019 (3 weeks).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

10.1.1.1. General Chair, Scientific Chair

- Laurent Amsaleg is general co-chair of ACM Intl. Conf. on Multimedia 2019.
- Guillaume Gravier is technical program co-chair for ACM Intl. Conf. on Multimedia 2019.

10.1.1.2. Member of the Organizing Committees

- Yannis Avrithis is workshops co-chair for ACM Intl. Conf. on Multimedia 2019.
- Simon Malinowski is member of the organizing commitee of the Advanced Analytics and Learning on Temporal Data workshop (Wurzburg, Sept. 2019)
- Vincent Claveau was Finance chair of SIGIR 2019, (Paris, July 2019)

10.1.2. Scientific Events: Selection

10.1.2.1. Chair of Conference Program Committees

- Vincent Claveau was co-chair of the text-mining workshop DeFT in Toulouse, July 2019

10.1.2.2. Member of the Conference Program Committees

- Laurent Amsaleg was a PC member of: Extraction et Gestion de Connaissances, ACM International Conference on Multimedia Retrieval, Multimedia Modeling, Content-Based Multimedia Indexing, IEEE International Conference on Multimedia & Expo, International Conference on Similarity Search and Applications, Multimedia Information Retrieval and Applications.
- Guillaume Gravier was PC member of: European Conf. on Information Retrieval, Intl. Symposium on Multimedia, ACM Intl. Conf. on Multimedia Retrieval, IEEE Workshop on Multimedia Modeling.

- Pascale Sébillot was a PC member of: Annual Meeting of the Association for Computational Linguistics (ACL), International Joint Conference on Artificial Intelligence on Artificial Intelligence (IJCAI), European Conference on Information Retrieval (ECIR), International Conference on MultiMedia Modeling (MMM), Terminologie et intelligence artificielle (TIA) workshop with TALN.
- Yannis Avrithis was PC member of: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), International Conference on Computer Vision (ICCV). He has been selected as Outstanding Reviewer of ICCV (91 selected out of 2506 reviewers in total).
- Vincent Claveau was a PC member of: European Conference on Information Retrieval (ECIR), EMNLP (demo), ACL (demo), TexMine workshop with EGC, DeFT workshop with TALN, TalMed workshop with MEDINFO, CORIA.
- Simon Malinowski is member of the program commitee of the Advanced Analytics and Learning on Temporal Data workshop (Wurzburg, Sept. 2019)
- Ewa Kijak was a PC member of: ACM Intl. Conf. on Multimedia (area chair), Intl. Conf. on Content-Based Multimedia Indexing (CBMI)
- Teddy Furon was a PC member of: IEEE Workshop on Information Forensics and Security (WIFS), Int. Conf. on Similarity Search and Applications (SISAP), IEEE Int. Conf. on Acoustic Speech and Signal Processing (ICASSP), ACM Information Hiding & Multimedia Security (IHMMSEC).

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- 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.
- Vincent Claveau is Chief Editor of the ISTE journal RIDoWS
- Vincent Claveau is member of the editorial board of the journal Traitement Automatique des Langues (TAL)

10.1.3.2. Reviewer - Reviewing Activities

- Pascale Sébillot reviewed for Traitement Automatique des Langues.
- Vincent Claveau was a reviewer for: Multimedia Tools and Applications, Technique et Sciences Informatiques, Documents numériques, Traitement Automatique des Langues, RIDoWS
- Ewa Kijak was a reviewer for: Multimedia Tools and Applications, Remote Sensing journal
- Laurent Amsaleg was a reviewer for: IEEE Transactions on Information Forensics and Security, ACM Transactions on Information Systems.
- Pascale Sébillot was a reviewer for: Conference Traitement Automatique des Langues Naturelles (TALN).
- Simon Malinowski was a reviewer for the Data Mining and Knowledge Discovery (DAMI) journal, Springer.
- Teddy Furon was a reviewer for: Entropy (MDPI), IEEE Trans. on Dependable and Secure Computing, IEEE Trans. on Information Forensics and Security, IEEE Trans. on Pattern Analysis and Machine Intelligence, IEEE Trans. on Image Processing.

10.1.4. Invited Talks

- Vincent Claveau gave an invited talk at "La vie sociale du fake sur les espaces numériques", Tours, France, nov 2019.
- Ewa Kijak gave a talk on "Tampering detection in images to fight fake news" for the GFAIH (Global Forum on Artificial Intelligence for Humanity) in Paris, during the workshop "AI and information disorder".

10.1.5. Leadership within the Scientific Community

- Laurent Amsaleg is a member of the Steering Committee of SISAP for the 2016-2020 term
- Laurent Amsaleg is a member of the Steering Committee of ACM Multimedia for the 2020-2023 term
- Guillaume Gravier is a member of the scientific board of the pre-GDR Traitement Automatique des Langues.
- Pascale Sébillot is a member of the permanent steering committee of Conf. francophone Traitement Automatique des Langues Naturelles.
- Pascale Sébillot is a member of the board of the pre-GDR Traitement Automatique des Langues, leader of the Intermodality and Multimodality working group.
- Vincent Claveau is deputy head of the GdR CNRS MaDICS https://www.madics.fr
- Vincent Claveau is the finance head of ARIA https://www.asso-aria.org

10.1.6. Scientific Expertise

- Laurent Amsaleg has been involved in the writing of the "Rapport de la mission sur les outils de reconnaissance des contenus protégés et les plateformes de partage en ligne : état de l'art et propositions. Conseil supérieur de la propriété littéraire et artistique, Hadopi".
- Guillaume Gravier was auditioned by the EC for the report "Towards European Media Sovereignty: An industrial media strategy to leverage data, algorithm and artificial intelligence" by Guillaume Klossa, special advisor to EC VP Andrus Ansip.
- Guillaume Gravier was appointed by Technical University of Delft as an evaluator for a professorship position.
- Teddy Furon is scientific advisor for the startup Lamark.

10.1.7. Research Administration

- Guillaume Gravier is deputy director of the Institut de Rercherche en Informatique et Systèmes Aléatoires (IRISA, UMR 6074).
- Guillaume Gravier is a member of the Board of the technology cluster Images & Réseaux.
- Pascale Sébillot was a member of the Conseil National des Universités 27th section (computer science) until Nov 2019.
- 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

Licence: Laurent Amsaleg & Teddy Furon, Indexation multimédia, 10h, L3-cours invité, ENS Rennes, France

Licence: Laurent Amsaleg, Bases de données avancées, 2h, L3-option génie mathématique, INSA 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

Licence: Pascale Sébillot, Natural Language Processing, 10h, L3, INSA Rennes, France

Master: Laurent Amsaleg, Bases de données avancées, 25h, M2, INSA Rennes, France

Master: Guillaume Gravier, Data analysis and probabilistic modeling, 30h, M2, University Rennes 1, France

Master: Pascale Sébillot, Natural Language Processing, 6h, M1, INSA Rennes, France

Master: Yannis Avrithis, Deep learning for vision, 20h, M2 SIF, France

Master: Yannis Avrithis, Computer vision, 30h, National and Kapodistrian University of Athens, Greece

Master : Simon Malinowski, Predictive analytics, 48h, M1 MIAGE, Univ. Rennes 1

Master : Simon Malinowski, Machine Learning, 32h, M2 MIAGE, Univ. Rennes 1

Master : Simon Malinowski, Symbolic data mining, 12h, M2 MIAGE, Univ. Rennes 1

Engineering school: Vincent Claveau, Machine Learning, 18h, 3rd year, INSA Rennes, France

Master: Vincent Claveau, Information Retrieval, 10h, M2 MIAGE, Univ. Rennes, France

Master: Ewa Kijak, Image processing, 55h, M1, ESIR, France

Master: Ewa Kijak, Supervised machine learning, 21h, M2R, University Rennes 1, France

Master: Ewa Kijak, Supervised machine learning, 15h, M2, University Rennes 1, France

Master: Ewa Kijak, Image classification, 45h, M1, ESIR, France

Master: Ewa Kijak, Image indexing, 17h, M2, University Rennes 1, France

Master: Ewa Kijak, Computer vision, 22h, M2, ESIR, France

Master: Teddy Furon, Rare Events, 20h, M2, Insa Rennes, France

10.2.2. Supervision

PhD in progress: Mathieu Laroze, Active learning on adaptive representations for object detection in high-resolution imaging, June 2016, Romain Dambreville, Chloe Friguet, Ewa Kijak & Sébastien Lefevre (with OBELIX, IRISA team)

PhD in progress: Hanwei Zhang, Deep Learning in Adversarial Contexts, October 2017, Laurent Amsaleg, Yannis Avrithis, Teddy Furon & Ewa Kijak

PhD in progress: Marzieh Gheisari Khorasgani, Secure identification in the Internet of Things, January 2018, Laurent Amsaleg & Teddy Furon

PhD in progress: Antoine Perquin, Universal speech synthesis through embeddings of massive heterogeneous data, October 2017, Laurent Amsaleg, Gwénolé Lecorvé & Damien Lolive (with Expression, IRISA team)

PhD in progress: Benoit Bonnet, Adversarial images, November 2019, Teddy Furon & Patrick Bas

PhD in progress: Tong Xue, Visualization and collaborative analysis of document collections and extracted knowledge for data journalism, October 2018, Laurent Amsaleg & Anastasia Bezerianos

PhD in progress: Mikail Demirdelen, User-adapted multi-document multimedia synthesis, started October 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 October 2017, Guillaume Gravier, Pascale Sébillot and François Goasdoué (with CEDAR, Inria team)

PhD in progress: Cyrielle Mallart, Incremental dynamic construction of knowledge graphs from text mining, started December 2018, Guillaume Gravier, Michel Le Nouy (Ouest-France), Pascale Sébillot

PhD in progress: François Torregrossa, Heterogeneous data embedding for professional search, started November 2018, Robin Allessiardo (So Local), Vincent Claveau, Guillaume Gravier

PhD in progress: Oriane Siméoni, Invariance and supervision in visual learning, started October 2016, Yannis Avrithis & Guillaume Gravier

PhD in progress: Yann Lifchitz, Few shot learning for object recognition in aerial images. Started March 2018, Yannis Avrithis & Sylvaine Picard (Safran Tech).

PhD in progress: Zhaohui Yang, learning visual models with minimal human supervision, started January 2018, Miaojing Shi & Yannis Avrithis & Chao Xu (Peking University, Beijing)

PhD in progress: Raquel Almeida, Learning hierarchichal models for multimedia data, started January 2019, Ewa Kijak & Simon Malinowski & Laurent Amsaleg

PhD defended: Cédric Maigrot, Détection de fausses informations dans les réseaux sociaux, defended April 2019, Laurent Amsaleg, Vincent Claveau & Ewa Kijak. See [1].

PhD defended: Ricardo Carlini-Sperandio, Unsupervised motif mining in multimedia time series, defended December 2019, Laurent Amsaleg.

10.2.3. Juries

- Guillaume Gravier was involved in the following juries:
  - PhD Eloi Zabrocki, Paris-Sorbonne Université, Nov. 2019, reviewer
  - HDR Benoît Favre, Aix-Marseille Université, Nov. 2019, reviewer
- Pascale Sébillot was involved in the following juries:
  - HDR Natalia Grabar, Université Paris-Saclay, May 2019, member
- Vincent Claveau was reviewer for the mid-term PhD auditions of: Ygor Gallina (LS2N), Tsanta Randriatsitohaina (LIMSI), Oussama Ahmia (IRISA), Hyun Jung Kang (MoDyCo)
- Vincent Claveau was a member of the PhD jury of C. Artaud (Univ. La Rochelle)
- Teddy Furon was reviewer for the mid-term PhD auditions of: Alexandre Sablayrolles (Inria Grenoble - Facebook FAIR), Solène Bernard (CRIStAL Lille)
- Teddy Furon was a member of the PhD jury of Clément Feutry (CentraleSupelec Paris).

10.3. Popularization

10.3.1. Articles and contents

- Vincent Claveau was interviewed for his activities about fake news detection in: hors-série 199 de Sciences et avenir Oct/Nov 2019; Sciences Ouest n375; Télérama; Science et vie; Science et avenir; 01net N 894.
- Vincent Claveau was interviewed for his activities about text mining the Grand Débat contributions: AEG infos; Ouest France.
- Ewa Kijak and Vincent Claveau were interviewed for the TV show Télématin Science, nov 2019.
- Ewa Kijak was interviewed about fake news and deepfake detection in: AFP, Le Télégramme, Data Analytics Post, Sciences Ouest n375, Le Temps, FranceTV Info (TV show "Vrai ou Fake"), tv5 monde
- Ewa Kijak was invited on the radio show "La méthode scientifique", France culture, on "Deepfake : faut-il le voir pour le croire ?"
- Teddy Furon was interviewed by Data Analytics Post about the connection of deepfake and watermarking.

10.3.2. Education

LINKMEDIA has been actively involved into the "L codent, L créent" initiative with Cyrielle Mallart and Oriane Siméoni.

10.3.3. Interventions

- Guillaume Gravier gave an invited talk entitled "Strengths, weaknesses and limits of AI illustrated" at the West Data Festival (professional forum, 100 attendees)
- Ewa Kijak gave an invited talk entitled "Fighting Fake news and deep fakes" at the CEPIC congress (Coordination of European Picture Agencies Stock, Press and Heritage)

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] C. MAIGROT.*Detection of misleading information on social networks*, Université de Rennes 1 [UR1], April 2019, https://tel.archives-ouvertes.fr/tel-02404234

Articles in International Peer-Reviewed Journal

[2] L. AMSALEG, M. E. HOULE, E. SCHUBERT.*Introduction to Special Issue of the 9th International Conference on Similarity Search and Applications (SISAP 2016)*, in "Information Systems", February 2019, vol. 80, 107 [*DOI :* 10.1016/J.IS.2018.11.006], https://hal.inria.fr/hal-02106914

[3] M. CUI, L. LI, M. SHI.*A Selective Biogeography-Based Optimizer Considering Resource Allocation for Large-Scale Global Optimization*, in "Computational Intelligence and Neuroscience", July 2019, p. 1-18 [*DOI :* 10.1155/2019/1240162], https://hal.inria.fr/hal-02383076

[4] C. DALLOUX, N. GRABAR, V. CLAVEAU.*Detecting negation: machine learning and a corpus for French*, in "Revue des Sciences et Technologies de l'Information - Série TSI : Technique et Science Informatiques", December 2019, p. 1-21, https://hal.archives-ouvertes.fr/hal-02402913

[5] S. LI, X. SONG, H. LU, L. ZENG, M. SHI, F. LIU.*Friend recommendation for cross marketing in online brand community based on intelligent attention allocation link prediction algorithm*, in "Expert Systems with Applications", July 2019, vol. 139, p. 1-11 [*DOI :* 10.1016/J.ESWA.2019.112839], https://hal.inria.fr/hal-02383107

[6] N. PAPANELOPOULOS, Y. AVRITHIS, S. KOLLIAS.*Revisiting the medial axis for planar shape decomposition*, in "Computer Vision and Image Understanding", February 2019, vol. 179, p. 66-78 [*DOI :* 10.1016/J.CVIU.2018.10.007], https://hal.inria.fr/hal-01930939

[7] O. SIMÉONI, A. ISCEN, G. TOLIAS, Y. AVRITHIS, O. CHUM.*Graph-based Particular Object Discovery*, in "Machine Vision and Applications", March 2019, vol. 30, n^{o} 2, p. 243-254 [*DOI :* 10.1007/S00138-019-01005-Z], https://hal.inria.fr/hal-02370238

International Conferences with Proceedings

[8] A. ANTONINI, F. BENATTI, E. KING, F. VIGNALE, G. GRAVIER.*Modelling changes in diaries, correspondence and authors' libraries to support research on reading: the READ-IT approach*, in "ODOCH 2019 - First International Workshop on Open Data and Ontologies for Cultural Heritage", Rome, Italy, June 2019, p. 1-12, https://hal.archives-ouvertes.fr/hal-02130008

[9] C. B. EL VAIGH, F. GOASDOUÉ, G. GRAVIER, P. SÉBILLOT.*Using Knowledge Base Semantics in Context-Aware Entity Linking*, in "DocEng 2019 - 19th ACM Symposium on Document Engineering", Berlin, Germany, ACM, September 2019, p. 1-10 [*DOI :* 10.1007/978-3-030-27520-4_8], https://hal.inria.fr/hal-02171981

[10] M. GHEISARI, T. FURON, L. AMSALEG.*Privacy Preserving Group Membership Verification and Identification*, in "CVPR 2019 - IEEE Conference on Computer Vision and Pattern Recognition", Long Beach, United States, IEEE, June 2019, p. 1-9, https://arxiv.org/abs/1904.10327 , https://hal.archives-ouvertes.fr/hal-02107442

[11] S. GÍSLASON, B. Þ. JÓNSSON, L. AMSALEG.*Integration of Exploration and Search: A Case Study of the M^3 Model*, in "MMM 2019 - 25th International Conference on MultiMedia Modeling", Thessaloniki, Greece, LNCS, Springer, December 2019, vol. 11295, p. 156-168 [*DOI :* 10.1007/978-3-030-05710-7_13], https://hal.inria.fr/hal-02106893

[12] A. ISCEN, G. TOLIAS, Y. AVRITHIS, O. CHUM.*Label Propagation for Deep Semi-supervised Learning*, in "CVPR 2019 - IEEE Computer Vision and Pattern Recognition Conference", Long Beach, United States, IEEE, June 2019, p. 1-10, https://hal.inria.fr/hal-02370297

[13] C. LEVERGER, S. MALINOWSKI, T. GUYET, V. LEMAIRE, A. BONDU, A. TERMIER.*Toward a Framework for Seasonal Time Series Forecasting Using Clustering*, in "IDEAL 2019", Manchester, United Kingdom, October 2019, p. 328-340 [*DOI :* 10.1007/978-3-030-33607-3_36], https://hal.inria.fr/hal-02371221

[14] Y. LIFCHITZ, Y. AVRITHIS, S. PICARD, A. BURSUC.*Dense Classification and Implanting for Few-Shot Learning*, in "CVPR 2019 - IEEE Computer Vision and Pattern Recognition Conference", Long Beach, United States, IEEE, June 2019, p. 1-10, https://hal.inria.fr/hal-02371279

[15] Y. LIU, M. SHI, Q. ZHAO, X. WANG.*Point in, Box out: Beyond Counting Persons in Crowds*, in "CVPR 2019 - IEEE Computer Vision and Pattern Recognition Conference", Long Beach, United States, IEEE, June 2019, p. 1-10, https://hal.inria.fr/hal-02383057

[16] H. RAGNARSDÓTTIR, Þ. ÞORLEIKSDÓTTIR, O. S. KHAN, B. Þ. JÓNSSON, G. Þ. GUÐMUNDSSON, J. ZAHÁLKA, S. RUDINAC, L. AMSALEG, M. WORRING.*Exquisitor: Breaking the Interaction Barrier for Exploration of 100 Million Images*, in "MM 2019 - 27th ACM International Conference on Multimedia", Nice, France, ACM, October 2019, p. 1029-1031 [*DOI :* 10.1145/3343031.3350580], https://hal.inria.fr/hal-02378272

[17] F. C. G. REIS, R. ALMEIDA, E. KIJAK, S. MALINOWSKI, S. J. F. GUIMARÃES, Z. DO PATROCINIO.*Combining convolutional side-outputs for road image segmentation*, in "IJCNN 2019 - International Joint Conference on Neural Networks", Budapest, Hungary, IEEE, July 2019, p. 1-8 [*DOI :* 10.1109/IJCNN.2019.8851843], https://hal.inria.fr/hal-02370834

[18] M. SHI, Z. YANG, C. XU, Q. CHEN.*Revisiting Perspective Information for Efficient Crowd Counting*, in "CVPR 2019 - IEEE Computer Vision and Pattern Recognition Conference", Long Beach, United States, IEEE, June 2019, p. 1-10, https://hal.inria.fr/hal-01831109

[19] O. SIMÉONI, Y. AVRITHIS, O. CHUM.*Local Features and Visual Words Emerge in Activations*, in "CVPR 2019 - IEEE Computer Vision and Pattern Recognition Conference", Long Beach, United States, IEEE, June 2019, p. 1-10, https://hal.archives-ouvertes.fr/hal-02374156

[20] R. SOUZA, R. ALMEIDA, R. MIRANDA, Z. KLEBER GONÇALVES DO PATROCINIO, S. MALINOWSKI, S. J. F. GUIMARÃES.*BRIEF-based mid-level representations for time series classification*, in "CIARP 2019 - 24th Iberoamerican Congress on Pattern Recognition", La Havane, Cuba, October 2019, p. 449-457 [*DOI :* 10.1007/978-3-030-33904-3_42], https://hal.inria.fr/hal-02371260

Conferences without Proceedings

[21] C. DALLOUX, V. CLAVEAU, N. GRABAR.*Speculation and negation detection in french biomedical corpora*, in "RANLP 2019 - Recent Advances in Natural Language Processing", Varna, Bulgaria, September 2019, p. 1-10, https://hal.archives-ouvertes.fr/hal-02284444

[22] T. FURON.*Watermarking error exponents in the presence of noise: The case of the dual hypercone detector*, in "IH&MMSEC'19 - 7th ACM Workshop on Information Hiding and Multimedia Security", Paris, France, ACM, July 2019, p. 173-181 [*DOI :* 10.1145/3335203.3335731], https://hal.archives-ouvertes.fr/hal-02122206

[23] M. GHEISARI, T. FURON, L. AMSALEG.*Group Membership Verification with Privacy: Sparse or Dense?*, in "WIFS 2019 - IEEE International Workshop on Information Forensics and Security", Delft, Netherlands, IEEE, December 2019, p. 1-6, https://hal.archives-ouvertes.fr/hal-02307926

[24] M. GHEISARI, T. FURON, L. AMSALEG, B. RAZEGHI, S. VOLOSHYNOVSKIY.*Aggregation and embedding for group membership verification*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, p. 2592-2596, https://arxiv.org/abs/1812.03943 - accepted at ICASSP 2019 [*DOI :* 10.1109/ICASSP.2019.8682422], https://hal.archives-ouvertes.fr/hal-02091036

[25] N. GRABAR, C. GROUIN, T. HAMON, V. CLAVEAU.*Annotated corpus with clinical cases in French*, in "TALN 2019 - 26e Conference on Traitement Automatique des Langues Naturelles", Toulouse, France, July 2019, p. 1-14, https://hal.archives-ouvertes.fr/hal-02391878

[26] N. GRABAR, C. GROUIN, T. HAMON, V. CLAVEAU.*Information Retrieval and Information Extraction from Clinical Cases. Presentation of the DEFT 2019 Challenge*, in "DEFT 2019 - Défi fouille de texte", Toulouse, France, July 2019, p. 1-10, https://hal.archives-ouvertes.fr/hal-02280852

[27] C. GROUIN, N. GRABAR, V. CLAVEAU, T. HAMON.*Clinical Case Reports for NLP*, in "BioNLP 2019 - 18th ACL Workshop on Biomedical Natural Language Processing", Florence, Italy, ACL, August 2019, p. 273–282 [*DOI :* 10.18653/v1/W19-5029], https://hal.archives-ouvertes.fr/hal-02371243

[28] A. PERQUIN, G. LECORVÉ, D. LOLIVE, L. AMSALEG.*Évaluation objective de plongements pour la synthèse de parole guidée par réseaux de neurones*, in "Traitement automatique du langage naturel", Toulouse, France, July 2019, https://hal.archives-ouvertes.fr/hal-02419483

[29] Y. WANG, R. EMONET, E. FROMONT, S. MALINOWSKI, E. MENAGER, L. MOSSER, R. TAVENARD.*Classification de séries temporelles basée sur des "shapelets" interprétables par réseaux de neurones antagonistes*, in "CAp 2019 - Conférence sur l'Apprentissage automatique", Toulouse, France, July 2019, p. 1-2, https://hal.archives-ouvertes.fr/hal-02268004

Scientific Books (or Scientific Book chapters)

[30] *Proceedings of the workshop DeFt - Défi Fouille de Textes*, July 2019, p. 1-97, https://hal.archives-ouvertes.fr/hal-02391768

[31] L. AMSALEG, O. CHELLY, M. E. HOULE, K.-I. KAWARABAYASHI, M. RADOVANOVIĆ, W. TREERATANAJARU.*Intrinsic Dimensionality Estimation within Tight Localities*, in "Proceedings of the 2019 SIAM International Conference on Data Mining", Society for Industrial and Applied Mathematics, May 2019, p. 181-189 [*DOI :* 10.1137/1.9781611975673.21], https://hal.inria.fr/hal-02125331

[32] C. FABRE, E. MORIN, S. ROSSET, P. SÉBILLOT.*Varia - Préface - 60-1*, ATALA, January 2020, vol. 60, n^{o} 1, p. 7-11, https://hal.archives-ouvertes.fr/hal-02433763

[33] E. MORIN, S. ROSSET, P. SÉBILLOT.*Varia - Préface - 59-1*, ATALA, May 2019, vol. 59, n^{o} 1, p. 7-11, https://hal.archives-ouvertes.fr/hal-01789046

Books or Proceedings Editing

[34] L. AMSALEG, B. HUET, M. LARSON, G. GRAVIER, H. HUNG, C.-W. NGO, W. T. OOI (editors). *Proceedings of the 27th ACM International Conference on Multimedia*, ACM Press, Nice, France, October 2019 [*DOI :* 10.1145/3343031], https://hal.inria.fr/hal-02378803

Research Reports

[35] B. CARAMIAUX, F. LOTTE, J. GEURTS, G. AMATO, M. BEHRMANN, F. BIMBOT, F. FALCHI, A. GARCIA, J. GIBERT, G. GRAVIER, H. HOLKEN, H. KOENITZ, S. LEFEBVRE, A. LIUTKUS, A. PERKIS, R. REDONDO, E. TURRIN, T. VIÉVILLE, E. VINCENT.*AI in the media and creative industries*, New European Media (NEM), April 2019, p. 1-35, https://arxiv.org/abs/1905.04175 , https://hal.inria.fr/hal-02125504

Other Publications

[36] A. ANTONINI, M. CARMEN SUÁREZ-FIGUEROA, A. ADAMOU, F. BENATTI, F. VIGNALE, G. GRAVIER, L. LUPI, B. OUVRY-VIAL.*Understanding the phenomenology of reading through modelling*, 2019, p. 1-22, Work in progress from the READ-IT program (www.readitproject.eu), Project Leader Brigitte Ouvry-Vial, funded by the EU- Joint Programming Initiative for Cultural Heritage, https://hal.archives-ouvertes.fr/hal-02305957

[37] A. ANTONINI, F. VIGNALE, G. GRAVIER, B. OUVRY-VIAL.*The Model of Reading : Modelling principles, Definitions, Schema, Alignments*, 2019, READ-IT Model of Reading -V2, https://hal-univ-lemans.archives-ouvertes.fr/hal-02301611

[38] A. ISCEN, G. TOLIAS, Y. AVRITHIS, O. CHUM.*Label Propagation for Deep Semi-supervised Learning*, November 2019, https://arxiv.org/abs/1904.04717 - Accepted to CVPR 2019, https://hal.inria.fr/hal-02370207

[39] A. ISCEN, G. TOLIAS, Y. AVRITHIS, O. CHUM, C. SCHMID.*Graph Convolutional Networks for Learning with Few Clean and many Noisy Labels*, November 2019, https://arxiv.org/abs/1910.00324 - working paper or preprint [*DOI :* 10.00324], https://hal.inria.fr/hal-02370212

[40] Y. LIFCHITZ, Y. AVRITHIS, S. PICARD, A. BURSUC.*Dense Classification and Implanting for Few-Shot Learning*, November 2019, https://arxiv.org/abs/1903.05050 - CVPR 2019, https://hal.inria.fr/hal-02370192

[41] O. SIMÉONI, Y. AVRITHIS, O. CHUM.*Local Features and Visual Words Emerge in Activations*, November 2019, https://arxiv.org/abs/1905.06358 - working paper or preprint, https://hal.inria.fr/hal-02370209

[42] O. SIMÉONI, M. BUDNIK, Y. AVRITHIS, G. GRAVIER.*Rethinking deep active learning: Using unlabeled data at model training*, November 2019, https://arxiv.org/abs/1911.08177 - working paper or preprint, https://hal.inria.fr/hal-02372102

[43] Z. YANG, M. SHI, Y. AVRITHIS, C. XU, V. FERRARI.*Training Object Detectors from Few Weakly-Labeled and Many Unlabeled Images*, December 2019, https://arxiv.org/abs/1912.00384 - working paper or preprint, https://hal.inria.fr/hal-02393688

[44] H. ZHANG, Y. AVRITHIS, T. FURON, L. AMSALEG.*Smooth Adversarial Examples*, November 2019, https://arxiv.org/abs/1903.11862 - working paper or preprint, https://hal.inria.fr/hal-02370202

[45] H. ZHANG, Y. AVRITHIS, T. FURON, L. AMSALEG.*Walking on the Edge: Fast, Low-Distortion Adversarial Examples*, December 2019, https://arxiv.org/abs/1912.02153 - 13 pages, 9 figures, https://hal.inria.fr/hal-02404216

References in notes

[46] L. AMSALEG, J. E. BAILEY, D. BARBE, S. ERFANI, M. E. HOULE, V. NGUYEN, M. RADOVANOVIĆ.*The Vulnerability of Learning to Adversarial Perturbation Increases with Intrinsic Dimensionality*, in "WIFS", 2017

[47] L. AMSALEG, O. CHELLY, T. FURON, S. GIRARD, M. E. HOULE, K.-I. KAWARABAYASHI, M. NETT.*Estimating Local Intrinsic Dimensionality*, in "KDD", 2015

[48] L. AMSALEG, G. Þ. GUÐMUNDSSON, B. Þ. JÓNSSON, M. J. FRANKLIN.*Prototyping a Web-Scale Multimedia Retrieval Service Using Spark*, in "ACM TOMCCAP", 2018, vol. 14, n^{o} 3s

[49] L. AMSALEG, B. Þ. JÓNSSON, H. LEJSEK.*Scalability of the NV-tree: Three Experiments*, in "SISAP", 2018

[50] R. BALU, T. FURON, L. AMSALEG.*Sketching techniques for very large matrix factorization*, in "ECIR", 2016

[51] S. BERRANI, H. BOUKADIDA, P. GROS.*Constraint Satisfaction Programming for Video Summarization*, in "ISM", 2013

[52] B. BIGGIO, F. ROLI.*Wild Patterns: Ten Years After the Rise of Adversarial Machine Learning*, in "Pattern Recognition", 2018

[53] P. BOSILJ.*Image indexing and retrieval using component trees*, Université de Bretagne Sud, 2016

[54] X. BOST.*A storytelling machine? : Automatic video summarization: the case of TV series*, University of Avignon, France, 2016

[55] M. BUDNIK, M. DEMIRDELEN, G. GRAVIER.*A Study on Multimodal Video Hyperlinking with Visual Aggregation*, in "ICME", 2018

[56] N. CARLINI, D. A. WAGNER.*Audio Adversarial Examples: Targeted Attacks on Speech-to-Text*, in "CoRR", 2018, vol. abs/1801.01944

[57] R. CARLINI SPERANDIO, S. MALINOWSKI, L. AMSALEG, R. TAVENARD.*Time Series Retrieval using DTW-Preserving Shapelets*, in "SISAP", 2018

[58] 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

[59] A. DELVINIOTI, H. JÉGOU, L. AMSALEG, M. E. HOULE.*Image Retrieval with Reciprocal and shared Nearest Neighbors*, in "VISAPP", 2014

[60] H. FARID.*Photo Forensics*, The MIT Press, 2016

[61] M. GAMBHIR, V. GUPTA.*Recent automatic text summarization techniques: a survey*, in "Artif. Intell. Rev.", 2017, vol. 47, n^{o} 1

[62] I. GOODFELLOW, Y. BENGIO, A. COURVILLE.*Deep Learning*, MIT Press, 2016

[63] G. GRAVIER, M. RAGOT, L. AMSALEG, R. BOIS, G. JADI, E. JAMET, L. MONCEAUX, P. SÉBILLOT.*Shaping-Up Multimedia Analytics: Needs and Expectations of Media Professionals*, in "MMM, Special Session Perspectives on Multimedia Analytics", 2016

[64] A. ISCEN, L. AMSALEG, T. FURON.*Scaling Group Testing Similarity Search*, in "ICMR", 2016

[65] A. ISCEN, G. TOLIAS, Y. AVRITHIS, O. CHUM.*Mining on Manifolds: Metric Learning without Labels*, in "CVPR", 2018

[66] B. Þ. JÓNSSON, G. TÓMASSON, H. SIGURÞÓRSSON, Á. ERÍKSDÓTTIR, L. AMSALEG, M. K. LARUSDOTTIR.*A Multi-Dimensional Data Model for Personal Photo Browsing*, in "MMM", 2015

[67] B. Þ. JÓNSSON, M. WORRING, J. ZAHÁLKA, S. RUDINAC, L. AMSALEG.*Ten Research Questions for Scalable Multimedia Analytics*, in "MMM, Special Session Perspectives on Multimedia Analytics", 2016

[68] H. KIM, P. GARRIDO, A. TEWARI, W. XU, J. THIES, N. NIESSNER, P. PÉREZ, C. RICHARDT, M. ZOLLHÖFER, C. THEOBALT.*Deep Video Portraits*, in "ACM TOG", 2018

[69] M. LAROZE, R. DAMBREVILLE, C. FRIGUET, E. KIJAK, S. LEFÈVRE.*Active Learning to Assist Annotation of Aerial Images in Environmental Surveys*, in "CBMI", 2018

[70] S. LEROUX, P. MOLCHANOV, P. SIMOENS, B. DHOEDT, T. BREUEL, J. KAUTZ.*IamNN: Iterative and Adaptive Mobile Neural Network for Efficient Image Classification*, in "CoRR", 2018, vol. abs/1804.10123

[71] A. LODS, S. MALINOWSKI, R. TAVENARD, L. AMSALEG.*Learning DTW-Preserving Shapelets*, in "IDA", 2017

[72] C. MAIGROT, E. KIJAK, V. CLAVEAU.*Context-Aware Forgery Localization in Social-Media Images: A Feature-Based Approach Evaluation*, in "ICIP", 2018

[73] D. SHAHAF, C. GUESTRIN.*Connecting the dots between news articles*, in "KDD", 2010

[74] M. SHI, H. CAESAR, V. FERRARI.*Weakly Supervised Object Localization Using Things and Stuff Transfer*, in "ICCV", 2017

[75] R. SICRE, Y. AVRITHIS, E. KIJAK, F. JURIE.*Unsupervised part learning for visual recognition*, in "CVPR", 2017

[76] R. SICRE, H. JÉGOU.*Memory Vectors for Particular Object Retrieval with Multiple Queries*, in "ICMR", 2015

[77] O. SIMÉONI, A. ISCEN, G. TOLIAS, Y. AVRITHIS, O. CHUM.*Unsupervised Object Discovery for Instance Recognition*, in "WACV", 2018

[78] O. SIMÉONI, A. ISCEN, G. TOLIAS, Y. AVRITHIS, O. CHUM.*Unsupervised Object Discovery for Instance Recognition*, in "WACV", 2018

[79] H. O. SONG, Y. XIANG, S. JEGELKA, S. SAVARESE.*Deep Metric Learning via Lifted Structured Feature Embedding*, in "CVPR", 2016

[80] C. TSAI, M. L. ALEXANDER, N. OKWARA, J. R. KENDER.*Highly Efficient Multimedia Event Recounting from User Semantic Preferences*, in "ICMR", 2014

[81] C. B. E. VAIGH, F. GOASDOUÉ, G. GRAVIER, P. SÉBILLOT.*Using Knowledge Base Semantics in Context-Aware Entity Linking*, in "DocEng", 2019

[82] O. VINYALS, A. TOSHEV, S. BENGIO, D. ERHAN.*Show and Tell: Lessons Learned from the 2015 MSCOCO Image Captioning Challenge*, in "TPAMI", 2017, vol. 39, n^{o} 4

[83] V. VUKOTIĆ, C. RAYMOND, G. GRAVIER.*Bidirectional Joint Representation Learning with Symmetrical Deep Neural Networks for Multimodal and Crossmodal Applications*, in "ICMR", 2016

[84] V. VUKOTIĆ, C. RAYMOND, G. GRAVIER.*Generative Adversarial Networks for Multimodal Representation Learning in Video Hyperlinking*, in "ICMR", 2017

[85] V. VUKOTIĆ.*Deep Neural Architectures for Automatic Representation Learning from Multimedia Multimodal Data*, INSA de Rennes, 2017

[86] J. WESTON, S. CHOPRA, A. BORDES.*Memory Networks*, in "CoRR", 2014, vol. abs/1410.3916

[87] H. YU, J. WANG, Z. HUANG, Y. YANG, W. XU.*Video Paragraph Captioning Using Hierarchical Recurrent Neural Networks*, in "CVPR", 2016

[88] J. ZAHÁLKA, M. WORRING.*Towards interactive, intelligent, and integrated multimedia analytics*, in "VAST", 2014

[89] L. ZHANG, M. SHI, Q. CHEN.*Crowd Counting via Scale-Adaptive Convolutional Neural Network*, in "WACV", 2018

[90] X. ZHANG, X. ZHOU, M. LIN, J. SUN.*ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices*, in "CoRR", 2017, vol. abs/1707.01083

[91] A. DA SILVA PINTO, D. MOREIRA, A. BHARATI, J. BROGAN, K. W. BOWYER, P. J. FLYNN, W. J. SCHEIRER, A. ROCHA.*Provenance filtering for multimedia phylogeny*, in "ICIP", 2017

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

Project-Team MIMETIC

Creation of the Team: 2011 January 01, updated into Project-Team: 2014 January 01

Keywords:

Computer Science and Digital Science:

A5.1.3. - Haptic interfaces
A5.1.5. - Body-based interfaces
A5.1.9. - User and perceptual studies
A5.4.2. - Activity recognition
A5.4.5. - Object tracking and motion analysis
A5.4.8. - Motion capture
A5.5.4. - Animation
A5.6. - Virtual reality, augmented reality
A5.6.1. - Virtual reality
A5.6.3. - Avatar simulation and embodiment
A5.6.4. - Multisensory feedback and interfaces
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
B9.4. - Sports

1. Team, Visitors, External Collaborators

Research Scientists

Franck Multon [Team leader, Inria, Senior Researcher, HDR]
Ludovic Hoyet [Inria, Researcher]

Faculty Members

Benoit Bideau [Univ Rennes II, Professor, HDR]
Nicolas Bideau [Univ Rennes II, Associate Professor]
Marc Christie [Univ Rennes I, Associate Professor]
Armel Crétual [Univ Rennes II, Associate Professor, HDR]
Georges Dumont [Ecole normale supérieure de Rennes, Professor, HDR]
Richard Kulpa [Univ Rennes II, Associate Professor, HDR]

Fabrice Lamarche [Univ Rennes I, Associate Professor]
Guillaume Nicolas [Univ Rennes II, Associate Professor]
Anne-Hélène Olivier [Univ Rennes II, Associate Professor]
Charles Pontonnier [Ecole normale supérieure de Rennes, Associate Professor, HDR]

External Collaborators

Charles Faure [Ecole normale supérieure de Rennes, PhD Student]
Carole Puil [IFPEK, Univ Rennes II, Enseignante à l'école de podologie, PhD Student]

Technical Staff

Ronan Gaugne [Univ Rennes I, Engineer, from Sep 2018, SED 10%]
Quentin Galvane [Inria, Engineer, until Sep 2019]
Anthony Mirabile [SATT Ouest Valorisation, Engineer, from Dec 2019]
Anthony Sorel [Univ Rennes II, Engineer]

PhD Students

Ludovic Burg [Univ Rennes I, from Oct 2019]
Jean Basset [Univ de Grenoble, from Oct 2018]
Rebecca Fribourg [Inria, from Sep 2017]
Diane Dewez [Inria, from oct 2018]
Florian Berton [Inria, from Nov 2017]
Hugo Brument [Inria, from Oct 2018]
Théo Perrin [ENS Rennes, Oct 2016 to Dec 2019]
Sebastien Cordillet [Univ Rennes II, Oct 2015 to Dec 2019]
Anabelle Limballe [Univ Rennes II, from Oct 2019]
Adèle Colas [Univ Rennes I, from Nov 2019]
Erwan Delhaye [Univ Rennes II, from Oct 2019]
Louise Demestre [Ecole normale supérieure de Rennes, from Oct 2019]
Olfa Haj Mahmoud [Cifre Faurecia]
Nils Hareng [Univ Rennes II, from Apr 2019]
Simon Hilt [Ecole normale supérieure de Rennes]
Claire Livet [Ecole normale supérieure de Rennes, from Sep 2019]
Amaury Louarn [Univ Rennes I]
Pierre Touzard [Univ Rennes II]
Benjamin Niay [Inria]
Nicolas Olivier [InterDigital, from Feb 2019]
Pierre Puchaud [Fondation Saint Cyr]
Xi Wang [Univ Rennes I]

Visiting Scientists

Michael Barnett Cowan [University of Waterloo, Professor, Sep 2019]
Sheryl Bourgaize [University Wilfrid Laurier, PhD student, from Apr 2019 until Jul 2019]
Robyn Grunberg [University Wilfrid Laurier, Master student, from Apr 2019 until Jul 2019]
Kristoffer Larsen Norheim [Aalborg university, PhD. Student, from Sep 2019 until Nov 2019]

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.

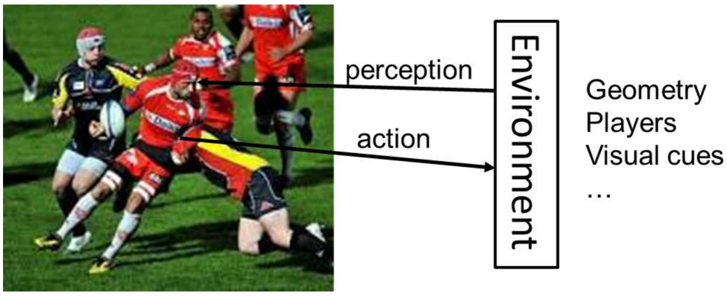


Figure 1. Main objective of MimeTIC: to better understand human activity in order to improve virtual human simulations. It involves modeling the complexity of human bodies, as well as of environments where to pick-up information and act upon.

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

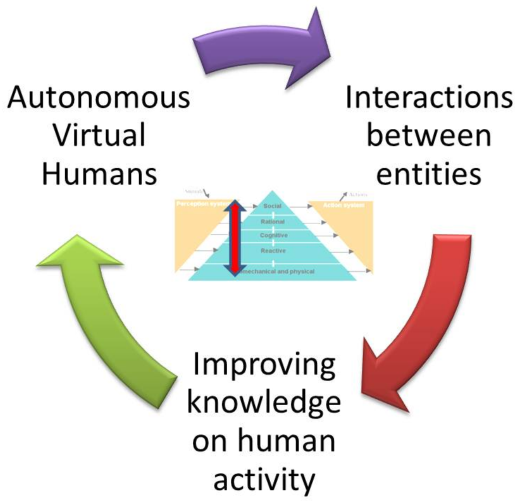


Figure 2. Research path of MimeTIC: coupling analysis and synthesis of human activity enables us to create more realistic autonomous characters and to evaluate assumptions about human motion control.

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.

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.

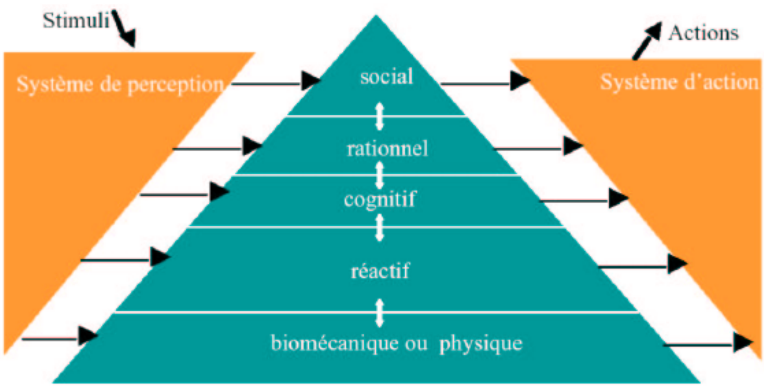


Figure 3. Layers of the motion control natural system in humans.

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. Computer animation

Computer animation is the branch of computer science devoted to models for the representation and simulation of the dynamic evolution of virtual environments. A first focus is the animation of virtual characters (behavior and motion). Through a deeper understanding of interactions using VR and through better perceptive, biomechanical and motion control models to simulate the evolution of dynamic systems, the Mimetic team has the ability to build more realistic, efficient and believable animations. Perceptual study also enables us to focus computation time on relevant information (i.e. leading to ensure natural motion from the perceptual points of view) and save time for unperceived details. The underlying challenges are (i) the computational efficiency of the system which needs to run in real-time in many situations, (ii) the capacity of the system to generalise/adapt to new situations for which data was not available or for which models were not defined for, and (iii) the variability of the models, i.e. their ability to handle many body morphologies and generate variations in motions that would be specific to each virtual character.

In many cases, however, these challenges cannot be addressed in isolation. Typically character behaviors also depend on the nature and the topology of the environment they are surrounded by. In essence, a character animation system should also rely on smarter representations of the environments, in order to better perceive the environment, and take contextualised decisions. Hence the animation of virtual characters in our context often requires to be coupled with models to represent the environment, reason, and plan both at a geometric level (can the character reach this location), and at a semantic level (should it use the sidewalk, the stairs, or the road). This represents the second focus. Underlying challenges are the ability to offer a compact, yet precise representation on which efficient path and motion planning can be performed, and on which high-level reasonning can be achieved.

Finally, a third scientific focus tied to the computer animation axis is digital storytelling. Evolved representations of motions and environments enable realistic animations. It is yet equally important to question how these event should be portrayed, when and under which angle. In essence, this means integrating *discourse models* into *story models*, the story representing the sequence of events which occur in a virtual environment, and the discourse representing how this story should be displayed (ie which events to show in which order and with which viewpoint). Underlying challenges are pertained to (i) narrative discourse representations, (ii) projections of the discourse into the geometry, planning camera trajectories and planning cuts between the viewpoints and (iii) means to interactively control the unfolding of the discourse.

By therefore establishing the foundations to build bridges between the high-level narrative structures, the semantic/geometric planning of motions and events, and low-level character animations, the Mimetic team adopts a principled and all-inclusive approach to the animation of virtual characters.

4. Application Domains

4.1. Animation, Autonomous Characters and Digital Storytelling

Computer Animation is one of the main application domain of the research work conducted in the MimeTIC team, in particular in relation to the entertainment and game industries. In these domains, creating virtual characters that are able to replicate real human motions and behaviours still highlights unanswered key challenges, especially as virtual characters are becoming more and required to populate virtual worlds. For instance, virtual characters are used to replace secondary actors and generate highly populated scenes that would be hard and costly to produce with real actors, which requires to create high quality replicas that appear, move and behave both individually and collectively like real humans. The three key challenges for the MimeTIC team are therefore (i) to create natural animations (i.e., virtual characters that move like real humans), (ii) to create autonomous characters (i.e., that behave like real humans) and (iii) to orchestrate the virtual characters so as to create interactive stories.

First, our challenge is therefore to create animations of virtual characters that are natural, in the largest sense of the term of moving like a real human real would. This challenge covers several aspects of Character Animation depending on the context of application, e.g., producing visually plausible or physically correct motions, producing natural motion sequences, etc. Our goal is therefore to develop novel methods for animating virtual characters, e.g., based on motion capture, data-driven approaches, or learning approaches. However, because of the complexity of human motion (e.g., the number of degrees of freedom that can be controled), resulting animations are not necessarily physically, biomechanically, or visually plaisible. For instance, current physics-based approaches produce physically correct motions but not necessarily perceptually plausible ones. All these reasons are why most entertainment industries still mainly rely on manual animation, e.g., in games and movies. Therefore, research in MimeTIC on character animation is also conducted with the goal of validating objective (e.g., physical, biomechanical) as well as subjective (e.g., visual plausibility) criteria.

Second, one of the main challenges in terms 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). For instance, autonomous characters play the role of actors that are driven by a scenario in video games and virtual storytelling. 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. In the MimeTIC team, our focus is therefore 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.

Finally, 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 computer-generated 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.

4.2. Fidelity of Virtual Reality

VR is a powerful tool for perception-action experiments. VR-based experimental platforms allow exposing a population to fully controlled stimuli that can be repeated from trial to trial with high accuracy. Factors can be isolated and objects manipulations (position, size, orientation, appearance, ..) are easy to perform. Stimuli can be interactive and adapted to participants' responses. Such interesting features allow researchers to use VR to perform experiments in sports, motion control, perceptual control laws, spatial cognition as well as person-person interactions. However, the interaction loop between users and their environment differs in virtual conditions in comparison with real conditions. When a user interact in an environment, movement from action and perception are closely related. While moving, the perceptual system (vision, proprioception,..) provides feedback about the users' own motion and information about the surrounding environment. That allows the user to adapt his/her trajectory to sudden changes in the environment and generate a safe and efficient motion. In virtual conditions, the interaction loop is more complex because it involves several material aspects.

First, the virtual environment is perceived through a numerical display which could affect the available information and thus could potentially introduce a bias. For example, studies observed a distance compression effect in VR, partially explained by the use of Head Mounted Display with reduced field of view and exerting a weight and torques on the user's head. Similarly, the perceived velocity in a VR environment differs from the real world velocity, introducing an additional bias. Other factors, such as the image contrast, delays in the displayed motion and the point of view can also influence efficiency in VR. The second point concerns the user's motion in the virtual world. The user can actually move if the virtual room is big enough or if wearing a head mounted display. Even with a real motion, authors showed that walking speed is decreased, personal space size is modified and navigation in VR is performed with increased gait instability. Although natural locomotion is certainly the most ecological approach, the physical limited size of VR setups prevents from using it most of the time. Locomotion interfaces are therefore required. Locomotion interfaces are made up of two components, a locomotion metaphor (device) and a transfer function (software), that can also introduce bias in the generated motion. Indeed, the actuating movement of the locomotion metaphor can significantly differ from real walking and the simulated motion depends on the transfer function applied. Locomotion interfaces cannot usually preserve all the sensory channels involved in locomotion.

When studying human behavior in VR, the aforementioned factors in the interaction loop potentially introduce bias both in the perception and in the generation of motor behavior trajectories. MimeTIC is working on the mandatory step of VR validation to make it usable for capturing and analyzing human motion.

4.3. 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: intra-variability 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), Qualisys, 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.

4.4. Sports

Sport is characterized by complex displacements and motions. One main objective is to understand the determinants of performance through the analysis of the motion itself. In the team, different sports have been studied such as the tennis serve, where the goal was to understand the contribution of each segment of the body in the performance but also the risk of injuries as well as other situation in cycling, swimming, fencing or soccer. Sports 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 sequentially 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 understand 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.

4.5. 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.

In real working condition, motion analysis and musculoskeletal risks assessment raise also many scientific and technological challenges. Similarly to virtual reality, fidelity of the working process may be affected by the measurement method. Wearing sensors or skin markers, together with the need of frequently calibrating the assessment system may change the way workers perform the tasks. Whatever the measurement is, classical ergonomic assessments generally address one specific parameter, such as posture, or force, or repetitions..., which makes it difficult to design a musculoskeletal risk factor that actually represent this risk. Another key scientific challenge is then to design new indicators that better capture the risk of musculoskeletal disorders. However, this indicator has to deal with the tradeoff between accurate biomechanical assessment and the difficulty to get reliable and the required information in real working conditions.

4.6. Locomotion and Interactions between walkers

Modeling and simulating locomotion and interactions between walkers is a very active, complex and competitive domain, interesting various disciplines such as mathematics, cognitive sciences, physics, computer graphics, rehabilitation etc. Locomotion and 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 should produce a locomotor movement while taking 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. MimeTIC is an international key contributor in the domain of understanding and simulating locomotion and interactions between walkers. By combining an approach based on Human Movement Sciences and Computer Sciences, the team focuses on locomotor invariants which characterize the generation of locomotor trajectories, conducts challenging experiments focusing on visuo-motor coordination involved during interactions between walkers both using real and virtual set-ups. One main challenge is to consider and model not only the "average" behaviour of healthy younger adult but also extend to specific populations considering the effect of pathology or the effect of age (kids, older adults). As a first example, 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. We extend this objective evaluation by using person-person interaction paradigm which allows studying visuo-motor strategies deficit in these specific populations.

Another fundamental question is the adaptation of the walking pattern according to anatomical constraints, such as pathologies in orthopedics, or adaptation to various human and non-human primates in paleoanthropoly. Hence, the question is to predict plausible locomotion according to a given morphology. This question raises fundamental questions about the variables that are regulated to control gait: balance control, minimum energy, minimum jerk...In MimeTIC we develop models and simulators to efficiently test hypothesis on gait control for given morphologies.

5. Highlights of the Year

5.1. Highlights of the Year

Members of the MimeTIC team / M2S laboratory carried out a PIA3 EUR (Ecole Universitaire de Recherche) project (DIGISPORT project) for the University of Rennes, which brings together the universities and Grandes Ecoles of the Rennes site. This project, with a total budget of €86 million, is funded by the Ministry of Higher Education, Research and Innovation to the tune of €5.9 million. The objective of DIGISPORT is to create a unique graduate school of international excellence in interdisciplinary training and research in digital sport sciences. This project aims to offer students in initial and continuing training an opportunity to build a study strategy suited to their professional goals and to the labor market. The digital revolution in sports and exercise is indeed already underway, at the confluence of the fast-growing markets of sport (€80 billion worldwide) and digital technology and connected objects (€207 billion worldwide). It leads to the emergence of new professions at the interface of these domains requiring skills in sports science, digital, electronics, and human and social sciences. Currently, education system is not designed to train this type of multi-skilled and agile students able to integrate an evolving labor market. DIGISPORT aims to link and structure training courses and research to promote a transversal approach uniting teaching and research staff around the new discipline of digital sport science and to address the new skills generated by the entry of sport into the digital age. The EUR will provide a coordinated training offer, from masters to doctoral level, that is resolutely interdisciplinary and strongly linked to research and innovation.
Based on previous scientific results in dynamic motion analysis, MimeTIC has developed an efficient software platform to carry-out biomechanical analysis based on motion capture data. "Customizable Toolbox for Musculoskeletal simulation" (CusToM) was delivered as an open source software available on a repository (https://github.com/anmuller/CusToM) and documented in [22]. CusToM is a MATLAB toolbox aiming at performing inverse dynamics-based musculoskeletal analyzes. This type of analysis is essential to access mechanical quantities of human motion in different fields such as clinic, ergonomics and sports. CusToM exhibits several features. It can generate a personalized musculoskeletal model, and can solve from motion capture data inverse kinematics, external forces estimation, inverse dynamics and muscle forces estimation

problems with a high level of customization for research purposes. It is also designed for non-expert users interested in motion analysis. CusToM is an OpenSource Software available with no restriction.

The Immersia VR platform has celebrated its 20 years of existence at Inria Rennes/IRISA center, within the "20ans d'Immersia" event (November 2019).

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: hal-01612939v1 - hal-01393066v1 - hal-01332716v1 - hal-01332711v2 - hal-01095084v1

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 automaticaly 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, Georges Dumont, Pierre Puchaud, Anthony Sorel, Claire Livet and Louise Demestre
- Contact: Charles Pontonnier
- Publications: Motion-based prediction of external forces and moments and back loading during manual material handling tasks - A case study with custom : a comparison of normal and altered gait with an ankle brace - Motion-based prediction of hands and feet contact efforts during asymmetric handling tasks - MusIC method enhancement by a sensitivity study of its performance: application to a lower limbs musculoskeletal model - Ground Reaction Forces and Moments Prediction of Challenging Motions: Fencing Lunges - CusToM: a Matlab toolbox for musculoskeletal simulation - The MusIC method: a fast and quasi-optimal solution to the muscle forces estimation problem

6.8. Immerstar Platform

Participants: Georges Dumont [contact], Ronan Gaugne, Anthony Sorel, Richard Kulpa.

With the two platforms of virtual reality, Immersia (http://www.irisa.fr/immersia/) and Immermove (http://m2slab.com/index.php/facilities-4/), 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 continued this year with 23 new cameras. 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 in 2017. In 2018, a Scale One haptic device has been installed. It allows, as in the CPER proposal, one or two handed haptic feedback in the full space covered by Immersia and possibility of carrying the user.

We celebrated the twentieth anniversary of the Immersia platform in November 2019 by inaugurating the new haptic equipment. We proposed scientific presentations and received 150 participants, and visits for support services where we received 50 persons.

7. New Results

7.1. Outline

In 2019, MimeTIC has maintained his activity in motion analysis, modelling and simulation, to support the idea that these approaches are strongly coupled in a motion analysis-synthesis loop. This idea has been applied to the main application domains of MimeTIC:

- Animation, autonomous characters and Digital Storytelling,
- Fidelity of Virtual Reality,
- Motion sensing of Human Activity,
- Sports,
- Ergonomics,
- and Locomotion and Interactions between walkers.

7.2. Animation, Autonomous Characters and Digital Storytelling

MimeTIC main research path consists in associating motion analysis and synthesis to enhance the naturalness in computer animation, with applications in movie previsualisation, and autonomous virtual character control. Thus, we pushed example-based techniques in order to reach a good tradeoff between simulation efficiency and naturalness of the results. In 2019, to achieve this goal, MimeTIC continued to explore the use of perceptual studies and model-based approaches, but also began to investigate deep learning, for example to control cameras in Movie previsualization.

7.2.1. VR as a Content Creation Tool for Movie Previsualisation

Participants: Marc Christie [contact], Quentin Galvane.

This work proposes a VR authoring system which provides intuitive ways of crafting visual sequences in 3D environments, both for expert animators and expert creatives. It is designed in mind to be applied animation and film industries, but can find broader applications (eg. in multimedia content creation). Creatives in animation and film productions have forever been exploring the use of new means to prototype their visual sequences before realizing them, by relying on hand-drawn storyboards, physical mockups or more recently 3D modelling and animation tools. However these 3D tools are designed in mind for dedicated animators rather than creatives such as film directors or directors of photography and remain complex to control and master. The proposed system is designed to reflect the traditional process through (i) a storyboarding mode that enables rapid creation of annotated still images, (ii) a previsualisation mode that enables the animation of the characters, objects and cameras, and (iii) a technical mode that enables the placement and animation of complex camera rigs (such as cameras cranes) and light rigs. Our methodology strongly relies on the benefits of VR manipulations to re-think how content creation can be performed in this specific context, typically how to animate contents in space and time. As a result, the proposed system is complimentary to existing tools, and provides a seamless back-and-forth process between all stages of previsualisation. We evaluated the tool with professional users to gather experts' perspectives on the specific benefits of VR in 3D content creation [36].

7.2.2. Deep Learning Techniques for Camera Trajectories

Participant: Marc Christie [contact].

Designing a camera motion controller which places and moves virtual cameras in relation with contents in a cinematographic way is a complex and challenging task. Many cinematographic rules exist, yet practice shows there are significant stylistic variations in how these can be applied. While contributions have attempted to encode rules by hand, this work is the very first to propose an end-to-end framework that automatically learns from real and synthetic movie sequences how the camera behaves in relation with contents. Our deep-learning framework extracts cinematic features of movies through a novel feature estimator trained on synthetic data, and learns camera behaviors from those extracted features, through the design of a Recurrent Neural Network (RNN) with a Mixture of Experts (MoE) gating mechanism. This cascaded network is designed to capture important variations in camera behaviors while ensuring the generalization capacity in the learning of similar behaviors. We demonstrate the features of our framework through experiments that highlight (i) the quality of our cinematic feature extractor (ii) the capacity to learn ranges of behaviors through the gating mechanism, and (iii) the ability to analyse the camera behaviors from a given input sequence, and automatically re-apply these behaviors on new virtual contents, offering exciting new possibilities towards a deeper understanding of cinematographic style and enhanced possibilities in transferring style from real to virtual. The work is a collaboration with the Beijing Film Academy in China.

7.2.3. Efficient Visibility Computation for Camera Control

Participants: Marc Christie [contact], Ludovic Burg.

Efficient visibility computation is a prominent requirement when designing automated camera control techniques for dynamic 3D environments; computer games, interactive storytelling or 3D media applications all need to track 3D entities while ensuring their visibility and delivering a smooth cinematographic experience. Addressing this problem requires to sample a very large set of potential camera positions and estimate visibility for each of them, which in practice is intractable. In this work, we introduce a novel technique to perform efficient visibility computation and anticipate occlusions. We first propose a GPU-rendering technique to sample visibility in Toric Space coordinates – a parametric space designed for camera control. We then rely on this visibility evaluation to compute an anticipation map which predicts the future visibility of a large set of cameras over a specified number of frames. We finally design a camera motion strategy that exploits this anticipation map to maximize the visibility of entities over time. The key features of our approach are demonstrated through comparison with classical ray-casting techniques on benchmark environments, and through an integration in multiple game-like 3D environment with heavy sparse and dense occluders.

7.2.4. Analysing and Predicting Inter-Observer Gaze Congruency

Participant: Marc Christie [contact].

In trying to better understand film media, we have been recently exploring the relation between the distribution of gaze states and the features of images, with the objective of establishing correlations to understand how films manipulate users gaze (and how gaze can be manipulated be re-editing film sequences). According to the literature regarding visual saliency, observers may exhibit considerable variations in their gaze behaviors. These variations are influenced by aspects such as cultural background, age or prior experiences, but also by features in the observed images. The dispersion between the gaze of different observers looking at the same image is commonly referred as inter-observer congruency (IOC). Predicting this congruence can be of great interest when it comes to study the visual perception of an image. We introduce a new method based on deep learning techniques to predict the IOC of an image [31]. This is achieved by first extracting features from an image through a deep convolutional network. We then show that using such features to train a model with a shallow network regression technique significantly improves the precision of the prediction over existing approaches.

7.2.5. Deep Saliency Models: the Quest for the Loss Function

Participant: Marc Christie [contact].

Following our idea of understanding gaze patterns in movie watching, and predicting these gaze patterns on sequences, we have been exploring the influence of loss functions in learning the visual saliency. Indeed, numerous models in the literature present new ways to design neural networks, to arrange gaze pattern data, or to extract as much high and low-level image features as possible in order to create the best saliency representation. However, one key part of a typical deep learning model is often neglected: the choice of the loss function. In this work, we explore some of the most popular loss functions that are used in deep saliency models [49]. We demonstrate that on a fixed network architecture, modifying the loss function can significantly improve (or depreciate) the results, hence emphasizing the importance of the choice of the loss function when designing a model. We also introduce new loss functions that have never been used for saliency prediction to our knowledge. And finally, we show that a linear combination of several well-chosen loss functions leads to significant improvements in performances on different datasets as well as on a different network architecture, hence demonstrating the robustness of a combined metric.

7.2.6. Contact Preserving Shape Transfer For Rigging-Free Motion Retargeting

Participants: Franck Multon [contact], Jean Basset.

In 2018, we introduced the idea of context graph to capture the relationship between body parts surfaces and enhance the quality of the motion retargetting problem. Hence, it becomes possible to retarget the motion of a source character to a target one while preserving the topological relationship between body parts surfaces. However this approach implies to strictly satisfy distance constraints between body parts, whereas some of them could be relaxed to preserve naturalness. In 2019, we introduced a new paradigm based on transfering the shape instead of encoding the pose constraints to tackle this problem [29].

Hence, retargeting a motion from a source to a target character is an important problem in computer animation, as it allows to reuse existing rigged databases or transfer motion capture to virtual characters. Surface based pose transfer is a promising approach to avoid the trial-and-error process when controlling the joint angles. The main contribution of this work is to investigate whether shape transfer instead of pose transfer would better preserve the original contextual meaning of the source pose. To this end, we propose an optimization-based method to deform the source shape+pose using three main energy functions: similarity to the target shape, body part volume preservation, and collision management (preserve existing contacts and prevent penetrations). The results show that our method is able to retarget complex poses, including several contacts, to very different morphologies. In particular, we introduce new contacts that are linked to the change in morphology, and which would be difficult to obtain with previous works based on pose transfer that aim at distance preservation between body parts. These preliminary results are encouraging and open several perspectives, such as decreasing computation time, and better understanding how to model pose and shape constraints.

7.2.7. The Influence of Step Length to Step Frequency Ratio on the Perception of Virtual Walking Motions

Participants: Ludovic Hoyet [contact], Benjamin Niay, Anne-Hélène Olivier.

Synthesizing walking motions that look realistic and diverse is a challenging task in animation, and even more when the target is to create realistic motions for large group of characters. Indeed, in order to keep a good trade-off between computational costs and realism, biomechanical constraints of human walk are not always fulfilled. In pilot experiments [38], [46], we have therefore started to investigate the ability of viewers to identify an invariant parameter of human walking named the walk ratio, representing the ratio between step length and step frequency of an individual, when applied to virtual humans. To this end, we recorded 4 actors (2 males, 2 females) walking at different freely chosen speeds, as well as at different combinations of step frequency and step length. We then performed pilot perceptual studies to identify the ability of viewers to detect the range of walk ratios considered as natural and compared it to the walk ratio freely chosen by the actor when performing walks at the same speeds. Our results will provide new considerations to drive the animation of walking virtual characters using the walk ratio as a parameter, which we believe could enable animators to control the speed of characters through simple parameters while retaining the naturalness of the locomotion.

7.3. Fidelity of Virtual Reality

MimeTIC wishes to promote the use of Virtual Reality to analyze and train human motor performance. It raises the fundamental question of the transfer of knowledge and skills acquired in VR to real life. In 2019, we put efforts in better understanding the potential fidelity of Virtual Reality experiences compared to real life experiences. It has been applied to various aspects of the interaction between pedestrians, but also the biomechanical fidelity of using haptic devices in highly constrainted conditions, such as hammering tasks.

7.3.1. Influence of Motion Speed on the Perception of Latency in Avatar Control

Participants: Ludovic Hoyet [contact], Richard Kulpa, Anthony Sorel, Franck Multon.

With the dissemination of Head Mounted Display devices in which users cannot see their body, simulating plausible avatars has become a key challenge. For fullbody interaction, avatar simulation and control involves several steps, such as capturing and processing the motion (or intentions) of the user using input interfaces, providing the resulting user state information to the simulation platform, computing a plausible adaptation of the virtual world, rendering the scene, and displaying the multisensory feedback to the user through output interfaces. All these steps imply that the displayed avatar motion appears to users with a delay (or latency) compared to their actual performance. Previous works have shown an impact of this delay on the perception-action loop, with possible impact on Presence and embodiment. We have explored [37] how the speed of the motion performed when controlling a fullbody avatar can impact the way people perceive and react to such a delay. We conducted an experiment where users were asked to follow a moving object with their finger, while embodied in a realistic avatar. We artificially increased the latency by introducing different levels of delays (up to 300ms) and measured their performance in the mentioned task, as well as their feeling about the perceived latency. Our results show that motion speed influenced the perception of latency: we found critical latencies of 80ms for medium and fast motion speeds, while the critical latency reached 120ms for a slow motion speed. We also noticed that performance is affected by both latency and motion speed, with higher speeds leading to decreased performance. Interestingly, we also found that performance was affected by latency before the critical latency for medium and fast speeds, but not for a slower speed. These findings could help to design immersive environments to minimize the effect of latency on the performance of the user, with potential impacts on Presence and embodiment.

7.3.2. Influence of Personality Traits and Body Awareness on the Sense of Embodiment in Virtual Reality

Participants: Ludovic Hoyet [contact], Rebecca Fribourg, Diane Dewez.

With the increasing use of avatars (i.e. the virtual representation of the user in a virtual environment) in virtual reality, it is important to identify the factors eliciting the sense of embodiment or the factors that can disrupt this feeling. This paper [35] reports an exploratory study aiming at identifying internal factors (personality traits and body awareness) that might cause either a resistance or a predisposition to feel a sense of embodiment towards a virtual avatar. To this purpose, we conducted an experiment (n=123) in which participants were immersed in a virtual environment and embodied in a gender-matched generic virtual avatar through a head-mounted display. After an exposure phase in which they had to perform a number of visuomotor tasks (during 2 minutes) a virtual character entered the virtual scene and stabbed the participants' virtual hand with a knife. The participants' sense of embodiment was measured, as well as several personality traits (Big Five traits and locus of control) and body awareness, to evaluate the influence of participants' personality on the acceptance of the virtual body. The major finding of the experiment is that the locus of control is linked to several components of embodiment: the sense of agency is positively correlated with an internal locus of control and the sense of body ownership is positively correlated with an external locus of control. Interestingly, both components are not influenced by the same traits, which confirms that they can appear independently. Taken together our results suggest that the locus of control could be a good predictor of the sense of embodiment when the user embodies an avatar with a similar physical appearance.

7.3.3. *Gaze Behaviour During Person-Person Interaction in VR*

Participants: Ludovic Hoyet, Anne-Hélène Olivier [contact], Florian Berton.

Figure 4. Conditions used in this work to understand the effect of VR setup on gaze beaviour in a collision avoidance task.

Simulating realistic interactions between virtual characters has been of interest to research communities for years, and is particularly important to automatically populate virtual environments. This problem requires to accurately understand and model how humans interact, which can be difficult to assess. In this context, Virtual Reality (VR) is a powerful tool to study human behaviour, especially as it allows assessing conditions which are both ecological and controlled. While VR was shown to allow realistic collision avoidance adaptations, in the frame of the ecological theory of perception and action, interactions between walkers can not solely be characterized through motion adaptations but also through the perception processes involved in such interactions. The objective of this study [30] is therefore to evaluate how different VR setups influence gaze behaviour during collision avoidance tasks between walkers. In collaboration with Julien Pettré in Rainbow team, we designed an experiment involving a collision avoidance task between a participant and another walker (real confederate or virtual character). During this task, we compared both the participant's locomotion and gaze behaviour in a real environment and the same situation in different VR setups (including a CAVE, a screen and a Head-Mounted Display) as illustrated on Figure4. Our results show that even if some quantitative differences exist, gaze behaviour is qualitatively similar between VR and real conditions. Especially, gaze behaviour in VR setups including a HMD is more in line with the real situation than the other setups. Furthermore, the outcome on motion adaptations confirms previous work, where collision avoidance behaviour is qualitatively similar in VR and real conditions. In conclusion, our results show that VR has potential for qualitative analysis of locomotion and gaze behaviour during collision avoidance. This opens perspectives in the design of new experiments to better understand human behaviour, in order to design more realistic virtual humans.

7.3.4. Gaze Anticipation in Curved Path in VR

Participants: Anne-Hélène Olivier [contact], Hugo Brument.

This work was performed in collaboration with Ferran Argelaguet-Sanz and Maud Marchal from Hybrid team [32]. We investigated whether the body anticipation synergies in real environments (REs) are preserved during navigation in virtual environments (VEs). Experimental studies related to the control of human locomotion in REs during curved trajectories report a top-down reorientation strategy with the reorientation of the gaze anticipating the reorientation of head, the shoulders and finally the global body motion. This anticipation behavior provides a stable reference frame to the walker to control and reorient whole-body according to the future direction. To assess body anticipation during navigation in VEs, we conducted an experiment where participants, wearing a head-mounted display, were asked to perform a lemniscate trajectory in a virtual environment (VE) using five different navigation techniques, including walking, virtual steering (hand, head or torso steering) and passive navigation. For the purpose of this experiment, we designed a new control law based on the powerlaw relation between speed and curvature during human walking. Taken together our results showed that a similar ordered top-down sequence of reorientation of the gaze, head and shoulders during curved trajectories between walking in REs and in VEs (for all the evaluated techniques). However, this anticipation mechanism significantly differs between physical walking in VE, where the anticipation is higher, and the other virtual navigation techniques. The results presented in this paper pave the way to the better understanding of the underlying mechanisms of human navigation in VEs and to the design of navigation techniques more adapted to humans

7.3.5. Validity of VR to Study Social Norms During Person-Person Interaction

Participants: Anne-Hélène Olivier [contact], Ludovic Hoyet, Florian Berton.

The modelling of virtual crowds for major events, such as the Oympics in Paris in 2024, takes into account the global proxemics standards of individuals without questioning the possible variability of these standards according to the space in which the interactions are performed. We know that body interactions (Goffman, 1974) are subject to rules whose variability is, at least in part, cultural (Hall, 1971). Obviously, these proxemics standards also address practical issues such as available space and space occupancy density. Our objective in this study was to understand the conditions which can explain that the discomfort felt and the adaptive behaviour performed differ when the interaction takes place in the same city and in spaces with identical occupancy densities. Especially, we focused on the effect of the social context of the environement. We aim at estimating the extent to which the prospect of attending a sports performance alters sensitivity to the transgression of proxemics norms. An additional objective was to evaluate whether virtual reality can help us to provide new insights in such a social context, where objective measures out-of-the lab are complex to perform. To answer this question, we designed in collaboration with Julien Pettré (Rainbow team) and colleagues in the field of sociology François Le Yondre, Théo Rougant and Tristan Duverne (Univ Rennes II) an experiment (in real context and then in virtual reality) in two different locations: a train station and the surroundings of a stadium before a league 1 football match) but with similar densities. The task performed by a confederate was to walk and stand excessively close to men aged 20 to 40. The individual's behaviour (not conscious of being a subject of the experiment) was observed by ethnography and explanatory interviews were conducted immediately afterwards. This same experiment was carried out in virtual reality conditions on the same type of population, modelling the two spaces and making it possible to acquire more precise and quantifiable data than in real conditions such as distances, travel time and eye fixations. The results show that the discomfort shown is much higher in the train station. The sporting context seems to participate in a form of relaxation of the norms of bodily interaction. Such a gap is not observable in virtual reality. From a methodological point of view, explicit interviews make it possible to usefully identify the reasons why virtual reality does not generate the same reactions, although it sometimes provokes the same sensitivity. Future work is needed to evaluate the effect of an increased immersion on such Social Science studies.

7.4. Motion Sensing of Human Activity

MimeTIC has a long experience in motion analysis in laboratory condition. In the MimeTIC project, we proposed to explore how these approaches could be transferred to ecological situations, with a lack of control on the experimental conditions. In the continuation of 2018, we have proposed to explore the use of cheap depth cameras solution for on-site motion analysis in ergonomics.

7.4.1. *Motion Analysis of Work Conditions Using Commercial Depth Cameras in Real Industrial Conditions*

Participant: Franck Multon [contact].

Based on a former PhD thesis (of Pierre Plantard) we have demonstrated the use of depth sensors in industry to assess risks of musculoskeletal disorders ar work. It has leaded to the creation of the the KIMEA software and of the Moovency start-up company in November 2018. In 2019 we published a synthesis work with new results [48] to demonstrate that such an approach can actually support the work of ergonomists in their goal to enhance the quality of life of workers in industry.

Hence, measuring human motion activity in real work condition is challenging as the environment is not controlled, while the worker should perform his/her task without perturbation. Since the early 2010s, affordable and easy-to-use depth cameras, such as the Microsoft Kinect system, have been applied for in-home entertainment for the general public. In this work, we evaluated such a system for the use in motion analysis in work conditions and propose software algorithms to enhance the tracking accuracy. Firstly, we highlighted the high performance of the system when used under the recommended setup without occlusions. However, when the position/orientation of the sensor changes, occlusions may occur and the performance of the system may decrease, making it difficult to be used in real work conditions. Secondly, we propose a software algorithm to adapt the system to challenging conditions with occlusions to enhance the robustness and accuracy. Thirdly, we show that real work condition assessment using such an adapted system leads to similar results comparing with those performed manually by ergonomists. These results show that such adapted systems could be used to support the ergonomists work by providing them with reproducible and objective information about the human movement. It consequently saves ergonomists time and effort and allows them to focus on high-level analysis and actions.

7.5. Sports

MimeTIC promotes the idea of coupling motion analysis and synthesis in various domains, especially sports. More specifically, we have a long experience and international leadership in using Virtual Reality for analyzing and training sports performance. In 2019, we continued to explore 1) how enhancing on-site sports motion analysis using models inspired from motion simulation techniques, and 2) how Virtual Reality could be used to analyze and train motor and perceptual skills in sports.

7.5.1. *Analysis of Fencing Lunge Accuracy and Response Time in Uncertain Conditions With an Innovative Simulator*

Participants: Anthony Sorel [contact], Richard Kulpa, Nicolas Bideau, Charles Pontonnier.

We conducted a study evaluating the motor control strategies implied by the introduction of uncertainty in the realization of lunge motions [27]. Lunge motion is one of the fundamental attacks used in modern fencing, asking for a high level of coordination, speed and accuracy to be efficient. The aim of the current paper was the assessment of fencer's performance and response time in lunge attacks under uncertain conditions. For this study, an innovative fencing lunge simulator was designed. The performance of 11 regional to national-level fencers performing lunges in Fixed, Moving and Uncertain conditions was assessed. The results highlighted notably that i) Accuracy and success decreased significantly in Moving and Uncertain conditions with regard to Fixed ones ii) Movement and Reaction times were also affected by the experimental conditions iii) Different fencer profiles were distinguishable among subjects. In conclusion, the hypothesis that fencers may privilege an adaptation to the attack conditions and preserve accuracy instead of privileging quickness was supported by the results. Such simulators may be further used to analyze in more detail the motor control strategies of fencers through the measure and processing of biomechanical quantities and a wider range of fencing levels.

It has also a great potential to be used as training device to improve fencer's performance to adapt his attack to controlled opponent's motion.

7.5.2. *Enactive Approach to Assess Perceived Speed Error during Walking and Running in Virtual Reality*

Participants: Théo Perrin, Richard Kulpa [contact], Charles Faure, Anthony Sorel, Benoit Bideau.

The recent development of virtual reality (VR) devices such as head mounted displays (HMDs) increases opportunities for applications at the confluence of physical activity and gaming. Recently, the fields of sport and fitness have turned to VR, including for locomotor activities, to enhance motor and energetic resources, as well as motivation and adherence. For example, VR can provide visual feedbacks during treadmill running, thereby reducing monotony and increasing the feeling of movement and engagement with the activity. However, the relevance of using VR tools during locomotion depends on the ability of these systems to provide natural immersive feelings, specifically a coherent perception of speed. The objective of this study is to estimate the error between actual and perceived locomotor speed in VE using an enactive approach, i.e. allowing an active control of the environment. Sixteen healthy individuals participated in the experiment, which consisted in walking and running on a motorized treadmill at speeds ranging from 3 to 11 km/h with 0.5 km/h increments, in a randomized order while wearing a HMD device (HTC Vive) displaying a virtual racetrack. Participants were instructed to match VE speed with what they perceived was their actual locomotion speed (LS), using a handheld Vive controller. They were able to modify the optic flow speed (OFS) with a 0.02 km/h increment/decrement accuracy. An optic flow multiplier (OFM) was computed based on the error between OFS and LS. It represents the gain that exists between the visually perceived speed and the real locomotion speed experienced by participants for each trial. For all conditions, the average of OFM was 1.00 $\pm.25$ to best match LS. This finding is at odds with previous works reporting an underestimation of speed perception in VR. It could be explained by the use of an enactive approach allowing an active and accurate matching of visually and proprioceptively perceived speeds by participants. But above all, our study showed that the perception of speed in VR is strongly individual, with some participants always overestimating and others constantly underestimating. Therefore, a general OFM should not be used to correct speed in VE to ensure congruence in speed perception, and we propose the use of individual models as recommendations for setting up locomotion-based VR applications.

7.5.3. *Acting Together, Acting Stronger? Interference Between Participants During Face-to-Face Cooperative Interception Task*

Participants: Charles Faure, Théo Perrin, Richard Kulpa [contact], Anthony Sorel, Anabelle Limballe, Benoit Bideau.

People generally coordinate their action to be more effective. However, in some cases, interference between them occur, resulting in an inefficient collaboration. The main goal of this study [16], [16] is to explore the way two persons regulate their actions when performing a cooperative task of ball interception, and how interference between them may occur. Starting face to face, twenty-four participants (twelve teams of two) had to physically intercept balls moving down from the roof to the floor in a virtual room. To this end, they controlled a virtual paddle attached to their hand moving along the anterior-posterior axis, and were not allowed to communicate. Results globally showed participants were often able to intercept balls without collision by dividing the interception space in two equivalent parts. However, an area of uncertainty (where many trials were not intercepted) appeared in the center of the scene highlighting the presence of interference between participants. The width of this area increased when situation became more complex and when less information was available. Moreover, participants often interpreted balls starting above them as balls they should intercept, even when these balls were in fine intercepted by their partner. Overall, results showed that team coordination emerges from between-participants interactions in this ball interception task and that interference between them depends on task complexity (uncertainty on partner's action and visual information available).

7.5.4. *Detection of Deceptive Motions in Rugby from Visual Motion Cues*

Participants: Richard Kulpa [contact], Anne-Hélène Olivier, Benoit Bideau.

Figure 5. Illustration of a rugby player interacting in VR with 3 representations of a virtual attacker.

Frequently, in rugby, players incorporate deceptive motions (e.g., a side-step) in order to pass their opponent. Previous works showed that expert defenders are more efficient in detecting deceptive motions. Performance was shown to be correlated with the evolution of the center of gravity of the attacker, suggesting that experts may rely on global motion cues. This study [19] aims at investigating whether a representation of center of gravity can be useful for training purposes, by using this representation alone or by combining it with the local motion cues given by body parts. We designed an experiment in virtual reality to control the motion cues available to the defenders. Sixteen healthy participants (seven experts and nine novices) acted as defenders while a virtual attacker approached. Participants completed two separate tasks. The first was a time occlusion perception task, occlusion after 100ms, 200ms or 300ms after the initial change in direction, thereafter participants indicated the passing direction of the attacker. The second was a perception-action task Figure 5, participants were instructed to intercept the oncoming attacker by displacing medio-laterally. The attacker performed either a non-deceptive motion, directly toward the final passing direction or a deceptive motion, initially toward a false direction before quickly reorienting to the true direction. There was a main effect of expertise, appearance, cut off times and motion on correct responses during both tasks. There was an interaction between visual appearance and expertise, and between motion type and expertise during the perception task, however, this interaction was not present during the perception-action task. We observed that experts maintained superiority in the perception of deceptive motion; however when the visual appearance is reduced to global motion alone the difference between novices and experts is reduced. We further explore the interactions and discuss the effects observed for the visual appearance and expertise.

7.5.5. IMU-based Motion Capture for Cycling Performance

Participants: Nicolas Bideau [contact], Guillaume Nicolas, Benoit Bideau, Sebastien Cordillet, Erwan Delhaye.

The quantification of 3D kinematical parameters such as body segment orientations and joint angles is important in the monitoring of cycling to provide relevant biomechanical parameters associated with performance optimization and/or injury prevention. Numerous experiments based on optoelectronic motion capture have been conducted in the laboratory to analyze kinematical variables (e.g., joint angles) during cycling. However, the assessment of kinematics in real conditions during training or competition is a challenging task, especially since conventional optoelectronic motion capture systems suffer from major drawbacks (restricted fields of view, cumbersome and time consuming) in this regard. To overcome these limitations, inertial measurement units (IMU) is a relevant solution for in situ cycling analysis as they allow a continuous data acquisition process throughout a cycling exercise. Beyond the common problem of the drift related to the integration of gyroscope data, one of the major issues in joint kinematics assessment using IMU devices lies is the misalignment of sensor axes with the anatomical body segment axis, which is not straightforward. Thus, we developed a novel sensor-to-segment calibration procedure for inertial sensor-based knee joint kinematics analysis during cycling. This procedure was designed to be feasible in-field, autonomously, and without any external operator or device. It combines a static standing up posture and a pedaling task. In comparison with conventional calibration methods commonly employed in gait analysis, the new method we proposed significantly improved the accuracy of 3D knee joint angle measurement when applied to cycling analysis [14]. As a second step related to the in-field application to track cycling, we estimated lower limb joint angles during a time trial on a velodrome. This integrative measurement exhibited the evolution of kinematic parameters in relation with distance but also with the track curvature [43].

7.6. Ergonomics

Ergonomics has become an important application domains in MimeTIC: being able to capture, analyze, and model human performance at work. In this domain, key challenge consists in using limited equipment to capture the physical activity of workers in real conditions. Hence, in 2019, we have designed a new approach to predict external forces using mainly motion capture data, and to personnalize the biomechanical capabilities (maximum feasible force/torque) of specific population.

7.6.1. Motion-based Prediction of External Forces

Participants: Charles Pontonnier [contact], Georges Dumont, Claire Livet, Anthony Sorel, Nicolas Bideau.

We proposed [21] a method to predict the external efforts exerted on a subject during handling tasks, only with a measure of his motion. These efforts are the contacts forces and moments on the ground and on the load carried by the subject. The method is based on a contact model initially developed to predict the ground reaction forces and moments. Discrete contact points are defined on the biomechanical model at the feet and the hands. An optimization technique computes the minimal forces at each of these points satisfying the dynamic equations of the biomechanical model and the load. The method was tested on a set of asymmetric handling tasks performed by 13 subjects and validated using force platforms and an instrumented load. For each task, predictions of the vertical forces obtained a RMSE of about 0.25 N/kg for the feet contacts and below 1 N/kg for the hands contacts. This method enables to quantitatively assess asymmetric handling tasks on the basis of kinetics variables without additional instrumentation such as force sensors and thus improve the ecological aspect of the studied tasks. We evaluated this method [23] on manual material handling (MMH) tasks. From a set of hypothesized contact points between the subject and the environment (ground and load), external forces were calculated as the minimal forces at each contact point while ensuring the dynamics equilibrium. Ground reaction forces and moments (GRF&M) and load contact forces and moments (LCF&M) were computed from motion data alone. With an inverse dynamics method, the predicted data were then used to compute kinetic variables such as back loading. On a cohort of 65 subjects performing MMH tasks, the mean correlation coefficients between predicted and experimentally measured GRF for the vertical, antero-posterior and medio-lateral components were 0.91 (0.08), 0.95 (0.03) and 0.94 (0.08), respectively. The associated RMSE were 0.51 N/kg, 0.22 N/kg and 0.19 N/kg. The correlation coefficient between L5/S1 joint moments computed from predicted and measured data was 0.95 with a RMSE of 14 Nm for the flexion / extension component. This method thus allows the assessment of MMH tasks without force platforms, which increases the ecological aspect of the tasks studied and enables performance of dynamic analyses in real settings outside the laboratory.

This method was successfully applied [24] on lunge motion that is a fundamental attack of modern fencing, asking for a high level of coordination, speed and accuracy. It consists in an explosive extension of the front leg accompanying an extension of the sword arm. In such motions, the direction of action and the way feet are oriented – guard position - are particularly challenging for a GRF&M prediction method. These methods are avalaible in CusToM software [22].

7.6.2. Biomechanics for Motion Analysis-Synthesis and Analysis of Torque Generation Capacities

Participants: Charles Pontonnier [contact], Georges Dumont, Nicolas Bideau, Guillaume Nicolas, Pierre Puchaud.

Characterization of muscle mechanism through the torque-angle and torque-velocity relationships [17] is critical for human movement evaluation and simulation. In-vivo determination of these relationships through dynamometric measurements and modelling is based on physiological and mathematical aspects. However, no investigation regarding the effects of the mathematical model and the physiological parameters underneath these models was found. The purpose of the current study was to compare the capacity of various torque-angle and torque-velocity models to fit experimental dynamometric measurement of the elbow and provide meaningful mechanical and physiological information. Therefore, varying mathematical function and physiological muscle parameters from the literature were tested. While a quadratic torque-angle model seemed to increase predicted to measured elbow torque fitting, a new power-based torque-velocity parametric model gave meaningful physiological values with similar fitting results to a classical torque-velocity model. This model is of interest to extract modelling and clinical knowledge characterizing the mechanical behavior the joint. Based on the same kind of methods, we proposed [25] to analyse torque generation capacities of a human knee. The torque generation capacities are often assessed for human performance, as well as for prediction of internal forces through musculoskeletal modelling. Scaling individual strength generation capacities is challenging but can provide physiologically meaningful perspectives. We propose to fit the models to isokinetic measurements of joint torques in different angle and angular velocity conditions. Assuming muscles are viscoelastic actuators, their entire architectures contribute to Joint Torque-Angle and Torque-Velocity Relationships (JTAR and JTVR respectively, and their coupling JTAVR) at the joint level. Experimental observation at different scales (muscle sarcomere, muscle fibre and joint) resulted in various JTAR models available in the literature. On the other side, JVTR models are often modelled without obvious physiological consistency. The above mentionned JTVR model was shown to increase physiological transparency of the elbow JTAVR. As those results might be joint-specific, we extended it to evaluate five JTAR and two JTVR models on the knee flexion and extension.

7.7. Locomotion and Interactions between Walkers

MimeTIC is a leader in the study and modeling of walkers' visuo-motor strategies. This implies to understand how humans generate their walking trajectories within an environment. This year, one main focus was to consider how the interaction models change with specific populations (including kids, older adults, concussed athletes or person on a wheelchair) as well as in specific environment (including narrow sidewalk, or environment with varying social context).

7.7.1. *Effect of Foot Stimulation on Locomotion*

Participants: Anne-Hélène Olivier, Armel Crétual [contact], Carole Puil.

Medio-Intern Element (EMI®) is a thin plantar insert used by podiatrists to treat postural deficiency. It was shown an influence of a 3 mm high EMI on Medio-Lateral (ML) displacement of the Centre of Pressure (CoP) of healthy participants in quasi-static standing. Recently it has been demonstrated that EMI has an impact on eyes vergence, and especially in population with plantar postural dysfunction. These effects were weakly assessed however and only using static tasks. Therefore, the objective of this work [53], [52], [41], was to evaluate the effect of the EMI while performing a locomotor task. We expected a controlateral deviation of the trajectory when this insert was located under one foot. Indeed, in previous studies dealing with bottom-up control of locomotion, it was shown that a 30 min podokinetic stimulation leads to a ML deviation of the trajectory when participants were asked to walk in a straight line with eyes closed. 20 healthy participants volunteered for this study. They participated into 3 different sessions in random order: either without EMI, with EMI under the right foot or under the left foot. Each session involved first, static tasks (with and without vision) to compare with previous work, then, dynamic locomotor tasks with 6 different conditions mixing trajectory (straight walking, 90° left or right turn) and vision (with and without vision) in random order. In static conditions, we computed the average ML position of the CoP. In dynamic conditions, we analyzed the difference in the final orientation of the locomotor trajectory with and without vision with an EMI with respect to this difference without the EMI. No significant effect of the EMI was observed for either static or dynamic conditions. Our results do not confirm the previous work in static conditions. Future work is needed to better understand the effect of this insert. In particular, our participants were healthy and it could be interesting to evaluate this effect in participants with postural deficiencies. These results would have an application in the design of new clinical tests.

7.7.2. Collision Avoidance between Walkers on a Curvilinear Path

Participants: Anne-Hélène Olivier [contact], Armel Crétual, Richard Kulpa, Anthony Sorel.

Crowded public spaces require humans to interact with what the environment affords to regulate interpersonal distance to avoid collisions. In the case of rectilinear trajectories, the collision avoidance behaviours have been extensively studied. It has been shown that the perceived action-opportunities of the walkers might be afforded based on a future distance of closest approach (also coined 'Minimal Predicted Distance', MPD). However, typical daily interactions do not always follow rectilinear but also curvilinear trajectories. In that context, it has been shown that a ball following a curvilinear trajectory can be successfully intercepted. However, it remains unclear whether the collision avoidance strategies in the well-studied linear trajectories can be transferred to curvilinear trajectories. Therefore, the aim of this work [44] was to examine collision avoidance behaviours when interacting with walkers following curvilinear trajectories. An experiment was designed using virtual reality in which 22 participants navigated toward a goal in a virtual environment with a joystick. A Virtual Human (VH) crossed the path of the participant from left and right with varying risks of collision. The VH followed either a curvilinear path with a fixed radius of 5 m or 10 m, approaching from in-front of and behind the participant, or a control rectilinear path. The final crossing distance, the number of collisions and inversions of initial crossing order were analysed to determine the success of the task. Further, MPD evolution over time and specific timing events was analysed across conditions. For a curvilinear path with a 5 m radius there were significantly more collisions when the VH approached from behind the participant, and significantly more inversions of the initial crossing order when the VH approached from in-front than the control rectilinear path. Final crossing distance was shorter when the VH followed a path with a 5 m radius from behind the participant. Finally, the evolution of the MPD over time was similar for paths with a 10 m radius when compared to the control rectilinear path, whereas the 5 m curvilinear paths had significant differences during the interaction. Overall, with few collisions and few inversions of crossing order we can conclude that participants were capable of interacting with virtual walkers on curvilinear trajectories. Further, the task was solved with similar avoidance adaptations to those observed for rectilinear interactions. However, paths with a smaller radius had more reported collisions and inversions. Future work should address how a curved trajectory during collision avoidance is perceived.

7.7.3. Collision Avoidance in Person-Specific Populations

Participants: Anne-Hélène Olivier [contact], Armel Crétual.

In the frame of the Inria BEAR associate team, we have used our 90° crossing paradigm to understand visuo-motor coordination in specific population. This is important, not only from a theoretical point of view but also to design more individual model of human locomotion in a dynamic environment.

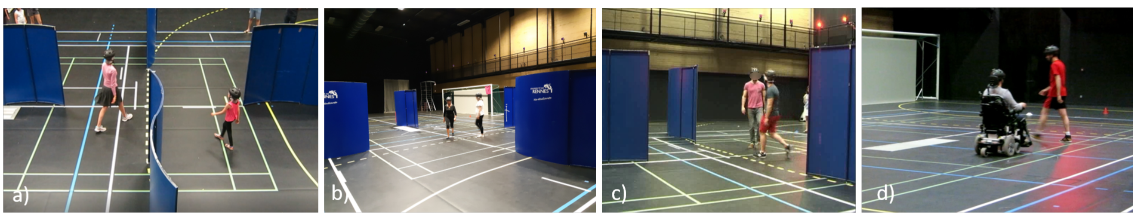


Figure 6. Illustration of person-person interaction experiments in a) kids, b) older adults, c) previously concussed athletes, d) a person on an electric powered wheelchair

We first investigated the effect of age on visuo-motor coordination by considering a collision avoidance task in kids (8-12 years) and older adults (65-74 years) as illustrated on Figure 6a,b. On one hand, middle-aged children have been shown to have poor perception-action coupling during static and dynamic collision avoidance tasks. Research has yet to examine whether perception-action coupling deficits persist in a dynamic collision avoidance task involving a child and another walker. In this work [26], [54], we invesigated whether the metric MPD(t) be used to examine collision avoidance strategies between children and adults. To this end, eighteen children (age: 10 ± 1.5 years) and eighteen adults (34 ± 9.6 years) walked while avoiding another participant (child or adult). Groups of three children and three adults were recruited per session. The results demonstrated that (1) MPD(t) can be used to predict future collisions in children, (2) MPD(t) is an absolute measure that is consistently lower when a child is involved compared to two adult walkers, (3) the individual passing second, even when it is a child, contributes more to MPD(t) than the walker passing first. It then appears that children have developed adult-like strategies during a collision avoidance task involving two walkers. Body anthropometrics should be considered when determining collision avoidance strategies between children and adults. On the other hand, every year, 1 in 3 older adults are likely to fall at least once and many falls occurs while walking where an individual needs to adapt to environmental hazards. Studies with older adults interacting within an environment showed difficulties in estimating time to arrival of vehicles, larger critical ratio and more variability in door aperture task as well as larger clearance distance when avoiding a moving object. The current study [51] aims to identify whether differences in collision avoidance behaviours of older adults during a person-person collision avoidance task are the result of age-related visuomotor processing deficits. Results showed that no collision occurred, where older and younger adults were able to act appropriately. However, larger threshold were needed to trigger avoidance when an older adult is second in crossing order, possibly due to visuomotor delays. Moreover, we observed more crossing inversions with older adults, which may suggest a poor visumotor processing. Finally, the clearance distance was samller when to older adults interact with each others, resulting in “risky” behaviours. Interestingly, social factors seems to be involved since when a young and an older adults interact, the young adult contribute more to solve the collision avoidance task.

In close relation with the Application Domain "Sports", we also investigated visuo-motor coordination during locomotion in previously concussed rugby-players (Figure 6c). Despite adherence to return-to-play guidelines, athletes with previous concussion exhibit persistent visuomotor deficits during static balance and visuomotor integration tasks such as collision avoidance months after returning to sport. Previous research in collision avoidance was done in a static setting, however less is known about visuomotor strategies utilized in dynamic scenarios, such as person-person interactions. In this context, during a collision avoidance locomotor task, individuals make adjustments to their path and/or their velocity in response to a risk of collision. These adjustments ensure that the clearance distance would be large enough such that no collision occurs. However, athletes with previous concussion may demonstrate impaired performance during a collision avoidance task requiring path adjustments based on visual information. The purpose of this study [55] was to investigate collision avoidance strategies when avoiding another walker between previously concussed athletes and healthy athletes. We hypothesized that previously concussed athletes would demonstrate altered trajectory adaptation and changes in individual contribution to the avoidance compared to healthy athletes. Preliminary results show that individuals with previous concussion demonstrated trajectory adaptation behaviours consistent with healthy athletes and young adults. However, previously concussed athletes passed with a reduced distance between themselves and the other walker when they are second in passage order at the crossing point. Athletes who have sustained a previous concussion show decreased collision avoidance behaviour. This behaviour results in a higher risk of a collision occurring, as individuals showed reduced contributions (i.e. creating physical space) to the avoidance of the collision. This change in typical behaviour on a visuomotor task may indicate a persistent deficit in perceptual abilities following concussion. Although trajectory adaptations were consistent with healthy athletes, these results suggest that athletes with previous concussion remain at an elevated risk of collision and possible injury following concussion recovery. This study provides novel insights and additional evidence that visuomotor and perceptual impairments persist following return to play in previously concussed athletes. Additionally, this protocol has important implications for the assessment and rehabilitation of visuomotor processes that are affected following a concussion. Future research could further develop this protocol to be used in sideline assessment, and guide treatment of concussions past clinical recovery.

7.7.4. Collision Avoidance between a Walker and an Electric Powered Wheelchair: Towards Smart Wheelchair

Participants: Anne-Hélène Olivier [contact], Armel Crétual.

In collaboration with Marie Babel and Julien Pettré from Inria Rainbow team, we are interested in the development of smart electric powered wheelchairs (EPW), which provide driver assistance. Developing smart assistance requires to better understand interactions between walkers and such vehicles. We focus on collision avoidance task between an EPW (fully operated by a human) and a walker, where the difference in the nature of the agents (weight, maximal speed, acceleration profiles) results into asymmetrical physical risk in case of a collision, for example due to the protection EPW provides to its driver, or the higher energy transferred to the walker during head-on collision. In this work [39], [47], our goal is to demonstrate that this physical risk asymmetry results into differences in the walker's behavior during collision avoidance in comparison to human-human situations. 20 participants (15 walkers and 5 EPW drivers) volunteered to this study. The experiment was performed in a 30mx20m gymnasium. We designed a collision avoidance task, where an EPW and a human walker moved towards a goal with orthogonal crossing trajectories (Figure6d). We recorded their trajectory among 246 trials (each trial being 1 collision avoidance). We compared the predicted passage order when they can first see each other with the one observed at the crossing point to identify if inversions occur during the interaction. Note that during walker-walker interactions it was shown that the initial passage order is almost systematically preserved all along the interaction up to the crossing point. We also computed the shape-to-shape clearance distance. We observed 23.7% of passage order inversion, specifically in 20.8% of trials where walkers were supposed to cross first, they crossed second. This means that walkers were more likely to pass behind the EPW than in front. On average, human walkers crossed first when having sufficient advance on the wheelchair to reach the crossing point. We estimated this advance up to 0.91m. The shape-to-shape clearance distance was influenced by the passage order at the crossing point, with larger distance when the walker cross first (M=0.78m) than second (M=0.34m). Results show that walkers set more conservative strategies when interacting with an EPW. By passing more frequently behind the EPW, they avoid risks of collisions that would lead to high energy transfer. Also, when they pass in front, they significantly increase the clearance distance, compared to cases where they pass behind. These results can then be linked to the difference in the physical characteristics of the walkers and EPW where asymmetry in the physical risks raised by collisions influence the strategies performed by the walkers in comparison with a similar walker-walker situation. This gives interesting insights in the task of modeling such interactions, indicating that geometrical terms are not sufficient to explain behaviours, physical terms linked to collision momentum should also be considered.

7.7.5. Collision Avoidance on a Narrow Sidewalk

Participant: Anne-Hélène Olivier [contact].

In the context of transportation research and a collaboration with the colleagues of Ifsttar (LEPSIS, LESCOT), we investigate person-person interaction when walking on a narrow sidewalk [34]. Narrow sidewalks are not the result of imagination nor a heritage of the former urban planning in the oldest cities. They exist in many modern cities, a simple web query provides a lot of examples in the world. In most cases, two pedestrians walking in opposite way cannot stay both on the sidewalk when they cross: one has to give a free way on the curb by stepping down on the road, which can generate risky situations for pedestrians. These situations are nowadays underestimated and so are the associated risk. In this context, driving simulators and walking simulators are useful tools to conduct studies in a safe environment with controlled conditions. Therefore, they can allow improving our knowledge on the way pedestrians interact on a narrow sidewalk and how drivers can react when facing this situation. This contribution aims to model the behaviours of simulated pedestrians, Non Player Characters (NPC). Using an interdisciplinary framework, we first identified from the literature psychosocial factors that should be involved in such interactions. Then, we designed a questionnaire to evaluate the impact of these factors on the perception of these interaction. Based on the main factors, we developed a perception model, and we modified the ORCA model, which is one of the most used for pedestrian collision avoidance simulation. Finally, we assessed the consistency of all our simulated interactions with a user study.

7.7.6. Shared Effort Model During Collision Avoidance

Participants: Anne-Hélène Olivier [contact], Armel Crétual.

In collaboration with Jose Grimaldo da Silva and Thierry Fraichard (Inria Grenoble), we finally designed a shared-effort model during interaction between a moving robot and a human relying on walker-walker collision avoidance data. [33]. Recent works in the domain of Human-Robot Motion (HRM) attempted to plan collision avoidance behavior that accounts for cooperation between agents. Cooperative collision avoidance between humans and robots should be conducted under several factors such as speed, heading and also human attention and intention. Based on some of these factors, people decide their crossing order during collision avoidance. However, whenever situations arise in which the choice crossing order is not consistent for people, the robot is forced to account for the possibility that both agents will assume the same role, a decision detrimental to collision avoidance. In our work we evaluate the boundary that separates the decision to avoid collision as first or last crosser. Approximating the uncertainty around this boundary allows our collision avoidance strategy to address this problem based on the insight that the robot should plan its collision avoidance motion in such a way that, even if agents, at first, incorrectly choose the same crossing order, they would be able to unambiguously perceive their crossing order on their following collision avoidance action.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Cifre Faurecia - Monitoring of gestual efficiency at work

Participants: Franck Multon [contact], Georges Dumont, Charles Pontonnier, Olfa Haj Mahmoud.

This Cifre contract has started in September 2018 for three years and is funding the PhD thesis of Olfa Haj Mamhoud. It consists in designing new methods based on depth cameras to monitor the activity of workers in production lines, compute the potential risk of musculoskeletal disorders, and efficiency compared to reference workers. It raises several fundamental questions, such as adapting previous methods to assess the risk of musculoskeletal disorders, as they generally rely on static poses whereas the worker is performing motion. Based on previous works in the team (previous Cifre PhD thesis of Pierre Plantard) we will provide 30Hz motion capture of the worker, that will enable us to evaluate various time-dependent assessment methods.

We will also explore how to estimate joint forces based and torques on such noisy and low-sampling motion data. We will then define a new assessment method based on these forces and torques.

The Cifre contracts funds the PhD salary and 10K€ per year for the supervision and management of the PhD thesis.

8.1.2. Cifre InterDigitial - Adaptive Avatar Customization for Immersive Experiences

Participants: Franck Multon [contact], Ludovic Hoyet, Nicolas Olivier.

This Cifre contract has started in February 2019 for three years and is funding the PhD thesis of Nicolas Olivier. The aim of the project is to design stylized avatars of users in immersive environment and digital arts such as videogames or cinema.

To this end, we will design a pipeline from motion and shape capture of the user to the simulation of the 3D real-time and stylized avatar. It will take hairs, eyes, face, body shape and motion into account. The key idea is to stylized both appaearance and motion to make avatar better correspond to the style of the movie of immersive experience. We will carry-out perceptual studies to better understand the expectation of the users when controlling stylized avatars, to maximize embodiment. The Cifre contracts funds the PhD salary and 15K€ per year for the supervision and management of the PhD thesis. This contract is also in collaboration with Hybrid team.

8.2. Bilateral Grants with Industry

8.2.1. Collaboration with company SolidAnim (Bordeaux, France)

Participants: Marc Christie [contact], Xi Wang.

Thsi contract started in November 2019 for three years. Its purpose is to explore novel means of performing depth detection for augmented reality applied to the film and broadcast industries. The grant serves to fund the PhD of Xi Wang.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Liv-Lab Breizh Digital Sport

Participants: Richard Kulpa [contact], Benoit Bideau, Franck Multon.

Our project aims, through new virtual reality and augmented reality technologies, to bring people who do not practice physical activity back into sport, whether for economic reasons, or for issues related to social and/or geographical isolation. To achieve this, the Brittany Region, accompanied by identified partners with complementary skills, proposes the development and networking of dedicated rooms at regional level. These existing rooms are chosen to be as close as possible to the target population, i.e. in priority areas: in the City Political District (QPV) for Rennes and Brest, and in the Rural Area to be Revitalized (ZRR) for Auray and Rostrenen. They will be redesigned to integrate these new technologies and attract target populations through the development of remote entertainment and collaborative applications. Indeed, the rooms will be connected to each other allowing participants to train together and create a community of practitioners. They will be equipped with simple sensors to evaluate their practices using a multidisciplinary cross-disciplinary approach, with biomechanical, physiological and psychological analyses (M2S/MimeTIC, CREAD and VIPS[2] laboratories). These evaluations will be used to propose physical activities that are progressive and adapted to the level of the practitioner. Access to objective data on their performance will be an additional motivating factor to keep these target audiences active. Support to local structures will allow them to extend their sporting experience after leaving Liv-Lab. Finally, subjects suffering from pathologies that are too disabling will be redirected to a health network, such as Rennes in the Living Lab ISAR (Innovation Santé Autonomie Rennes).

9.2. National Initiatives

9.2.1. ANR

9.2.1.1. ANR PRCE Cineviz

Participants: Marc Christie [contact], Quentin Galvane.

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. ANR PRC Capacities

Participants: Charles Pontonnier [contact], Georges Dumont, Pierre Puchaud, Claire Livet, Anthony Sorel.

This project is leaded by Christophe Sauret, from INI/CERAH. The project objective is to build a series of biomechanical indices characterizing the biomechanical difficulty for a wide range of urban environmental situations. These indices will rely on different biomechanical parameters such as proximity to joint limits, forces applied on the handrims, mechanical work, muscle and articular stresses, etc. The definition of a more comprehensive index, called Comprehensive BioMechanical (CBM) cost, including several of the previous indices, will also be a challenging objective. The results of this project would then be used in the first place in VALMOBILE application to assist MWC users in selecting optimal route in Valenciennes agglomeration (project founded by the French National Agency for Urban Renewal and the North Department of France). The MimeTIC team is involved on the musculoskeletal simulation issues and the biomechanical costs definition.

9.2.1.3. ANR JCJC Per2

Participants: Ludovic Hoyet [contact], Benjamin Niay, Anne-Hélène Olivier, Richard Kulpa, Franck Multon.

Per2 is a 42-month ANR JCJC project (2018-2022) entitled *Perception-based Human Motion Personalisation* (Budget: 280kE; website: https://project.inria.fr/per2/)

The objective of this project is to focus on how viewers perceive motion variations to automatically produce natural motion personalisation accounting for inter-individual variations. In short, our goal is to automate the creation of motion variations to represent given individuals according to their own characteristics, and to produce natural variations that are perceived and identified as such by users. Challenges addressed in this project consist in (i) understanding and quantifying what makes motions of individuals perceptually different, (ii) synthesising motion variations based on these identified relevant perceptual features, according to given individual characteristics, and (iii) leveraging even further the synthesis of motion variations and to explore their creation for interactive large-scale scenarios where both performance and realism are critical.

This work is performed in collaboration with Julien Pettré from Rainbow team.

9.2.1.4. ANR PRCI HoBis

Participants: Franck Multon [contact], Armel Crétual, Georges Dumont, Charles Pontonnier, Anthony Sorel.

Hobis is a 42-month ANR collaborative (PRCI) project (2018-2022) entitled *Hominin BipedalismS: Exploration of bipedal gaits in Hominins thanks to Specimen-Specific Functional Morphology*. HoBis is leaded by the Museum Nationale d'Histoires Naturelles (CNRS), with CNRS/LAAS, and Antwerpen University (Belgium), with a total of 541KE budget (140KE for MimeTIC).

HoBiS (Hominin BipedalismS) is a pluridisciplinary research project, fundamental in nature and centred on palaeoanthropological questions related to habitual bipedalism, one of the most striking features of the human lineage. Recent discoveries (up to 7 My) highlight an unexpected diversity of locomotor anatomies in Hominins that lead palaeoanthropologists to hypothesize that habitual bipedal locomotion took distinct shapes through our phylogenetic history. In early Hominins, this diversity could reveal a high degree of locomotor plasticity which favoured their evolutionary success in the changing environments of the late Miocene and Pliocene. Furthermore, one can hypothesize based on biomechanical theory that differences in gait characteristics, even slight, have impacted the energy balance of hominin species and thus their evolutionary success. However, given the fragmented nature of fossil specimens , previous morphometric and anatomo-functional approaches developed by biologists and palaeoanthropologists, do not allow the assessment of the biomechanical and energetic impacts of such subtle morphological differences, and the manners in which hominin species walked still remains unknown. To tackle this problem, HoBiS proposes as main objective a totally new specimen-specific approach in evolutionary anthropology named Specimen-Specific Functional Morphology: inferring plausible complete locomotor anatomies based on fossil remains, to link these reconstructed anatomies and corresponding musculoskeletal models (MSM) with plausible gaits using simulations. Both sub-objectives will make use of an extensive comparative anatomical and gait biomechanical data bases (challenges). To this end, we will integrate anatomical and functional studies, tools for anatomical modelling, optimization and simulation rooted in informatics, biomechanics, and robotics, to build an in-silico decision-support system (DSS). This DSS will provide biomechanical simulations and energetic estimations of the most plausible bipedal gaits for a variety of hominin species based on available remains, from partial to well-preserved specimens. To achieve this main objective, the project will address the following sub-objectives and challenges

MimeTIC is Leader of WP3 "Biomechanical simulation", aiming at predicting plausible bipedal locomotion based on paleoanthropological heuristics and a given MSM.

9.2.1.5. Labex CominLabs : Moonlight

Participants: Guillaume Nicolas [contact], Nicolas Bideau.

Moonlight is a 2-year Labex Cominlabs project (2018-2019). Amount: 55kE (including a one-year postdoctoral fellowship). Partners: Granit Team IRISA (http://www-granit.irisa.fr/fr/), M2S Lab.

The Moonlight project is part of an effort to transpose the tools and methodologies used in motion capture from optoelectronic equipment to inertial unit devices. More specifically, the overall objective of Moonlight project is to design a new embedded system in order to analyze cyclists' movements in real conditions, i.e. outside of the laboratory. This requires to estimate reliable 3D joint angles, lower limb kinematics and pedals orientation. IMUs are used as an alternative to optoelectronical motion capture but some challenges have to be addressed as regards to sensor-to-segment misalignment and drift. Indeed, a real time accurate orientation of the crank is necessary to get limb position. To achieve this goal, data fusion algorithms between IMU data and pedal orientation are implemented. A wireless sensor network with accurate time synchronization mechanism is needed to process data fusion from all sensor's nodes on a tablet. Finally, the system deals with size, energy consumption and ease-to-use constraints.

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.

The project funded by IFCE ended in 2018 but we got a 30K€ budget from SATT Ouest Valorisation in order to finish the development of the technological prototype, and to evaluate the possibility to patent the process, and transfer it to private companies. This project is in collaboration with LEGO lab. in University Bretagne Sud, and CAIRN Inria team.

9.2.2.2. French Federation of Tennis

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 training load.

9.2.3. Chaire Safran-Saint-Cyr "the enhanced soldier in the digital battlefield"

Participants: Charles Pontonnier [contact], Pierre Puchaud.

The chaire has the goal to answer to scientific questions accompanying the evolution of the technologies equipping the soldiers in mission. In this scheme, the MimeTIC team is involved in generic and specific musculoskeletal models for the prototyping of load carriage assistive devices (exoskeletons). Chair sponsored by SAFRAN group, led by Yvon Erhel (Professor, Ecoles de Sainr-Cyr Coëtquidan).

9.2.4. AUTOMA-PIED

Participants: Anne-Hélène Olivier [contact], Armel Crétual, Anthony Sorel.

The AUTOMA-PIED project is driven by IFSTTAR. Using a set-up in virtual reality, the first objective of the project aims at comparing pedestrian behaviour (young and older adults) when interacting with traditional or autonomous vehicles in a street crossing scenario. The second objective is to identify postural cues that can predict whether or not the pedestrian is about to cross the street.

9.2.5. IPL Avatar

Participants: Ludovic Hoyet [contact], Franck Multon.

This project aims at design avatars (i.e., the user's representation in virtual environments) that are better embodied, more interactive and more social, through improving all the pipeline related to avatars, from acquisition and simulation, to designing novel interaction paradigms and multi-sensory feedback. It involves 6 Inria teams (GraphDeco, Hybrid, Loki, MimeTIC, Morpheo, Potioc), Prof. Mel Slater (Uni. Barcelona), and 2 industrial partners (InterDigitak and Faurecia).

Website: http://avatar.inria.fr

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

9.3.1.1. H2020 ICT-25 PRESENT

Participants: Marc Christie, Ludovic Hoyet [contact], Anne-Hélène Olivier, Alberto Jovane, Adèle Colas.

This European project aims at creating virtual characters that are realistic in looks and behaviour, and who can act as trustworthy guardians and guides in the interfaces for AR, VR and more traditional forms of media. It is conducted in collaboration with industrial partners The Framestore Ltd, Cubic Motion Ltd, InfoCert Spa, Brainstorm Multimedia S.L., Creative Workers - Creatieve Werkers VZW, and academic partners Universidad Pompeu Fabra and Universität Augsburg.

9.3.1.2. JPI-CH SCHEDAR

Participants: Franck Multon [contact], Richard Kulpa.

SCHEDAR (Safeguarding the Cultural HEritage of Dance through Augmented Reality) is a Joint Program Initiative for preserving immaterial cultural heritage. The project started in June 2018 and will finish December 2021. It is coordinated by University of Cyprus, in collaboration with Algolysis LTD (Cyprus), University of Warwick (UK), University of Reims Champagne Ardennes (France).

Dance is an integral part of any culture. Through its choreography and costumes dance imparts richness and uniqueness to that culture. Over the last decade, technological developments have been exploited to record, curate, remediate, provide access, preserve and protect tangible CH. However, intangible assets, such as dance, has largely been excluded from this previous work. Recent computing advances have enabled the accurate 3D digitization of human motion. Such systems provide a new means for capturing, preserving and subsequently re-creating ICH which goes far beyond traditional written or imaging approaches. However, 3D motion data is expensive to create and maintain, encompassed semantic information is difficult to extract and formulate, and current software tools to search and visualize this data are too complex for most end-users. SCHEDAR will provide novel solutions to the three key challenges of archiving, re-using and re-purposing, and ultimately disseminating ICH motion data. In addition, we will devise a comprehensive set of new guidelines, a framework and software tools for leveraging existing ICH motion databases. Data acquisition will be undertaken holistically; encompassing data related to the performance, the performer, the kind of the dance, the hidden/untold story, etc. Innovative use of state-of-the-art multisensory Augmented Reality technology will enable direct interaction with the dance, providing new experiences and training in traditional dance which is key to ensure this rich culture asset is preserved for future generations. MimeTIC is responsible for WP3 "Dance Data Enhancement".

9.4. International Initiatives

9.4.1. Inria Associate Teams Not Involved in an Inria International Labs

9.4.1.1. BEAR

Title: from BEhavioral Analysis to modeling and simulation of interactions between walkeRs

International Partner: Michael Cinelli (Wilfrid Laurier University, Canada) and Michael Barnett Cowann (University of Waterloo, Canada)

Start year: 2019

See also: https://sites.google.com/view/inriabearproject/home

Interactions between individuals 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 or in more dynamical and strategic situations such as sports motion, we take in information about our surrounding environment in order to interact with people, move without collision, alone or in a group, intercept, meet or escape other people. In this context, the BEAR project is a collaboration between researchers from Inria Rennes (Computer Sciences) and Waterloo universities (Kinesiology-Neuroscience). The project aims at providing more realistic models and simulations of interactions between pedestrians, for various applications such as rehabilitation, computer graphics, or robotics. The originality of the project is to investigate the complexity of human interactions from a human motor control perspective, considering the strong coupling between pedestrians' visual perception and their locomotor adaptations. We will investigate how people gather the relevant information to control their motion. To provide generic models considering the inter-individual variability of humans, we will consider both normal populations and specific populations (children, older adults, injured, diseased ...) for whom an altered perception can modify their motion. The strength of this project is the complementarity of theinvolved teams. While all researchers will equally perform experiments on interactions between pedestrians, the researchers from Waterloo will take the lead to identify the relevant behavioral variables that will be used mainly by the researchers from Rennes to design the new models and simulations.

9.4.2. International Mobility Grant

- Mitacs Globalink grant: Perception-Action Integration in Collision Avoidance in Older Adults, Robyn Grundberg, University Wilfrid Laurier, Canada (April-July 2019)
- Mitacs Globalink grant: Influence of walking speed and trunk sway on collision avoidance with a virtual human, Sheryl Bourgaize, University Wilfrid Laurier, Canada (April-July 2019).

9.4.3. Inria International Partners

9.4.3.1. Informal International Partners

- Dr. Rachel McDonnell, Trinity College Dublin, Ireland (on-going collaboration with Ludovic Hoyet)
- Prof. Carol O Sullivan, Trinity College Dublin, Ireland (on-going collaboration with Ludovic Hoyet)
- Prof Michael Cinelli, University Wilfrid Laurier, Waterloo, Canada (on-going collaboration with Anne-Hélène Olivier)
- Prof Michael Barnet-Cowann, University of Waterloo, Waterloo, Canada (on-going collaboration with Anne-Hélène Olivier)
- Prof. Hui Huang, Shenzhen University (on-going collaboration with Marc Christie)
- Prof. Baoquan Chen, Pekin University (on-going collaboration with Marc Christie)
- Dr. Bin Wang, Beijing Film Academy University (on-going collaboration with Marc Christie)

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- Michael Barnett-Cowan (from Waterloo University, Canada): Visiting Professor, Research Chair of America, Rennes 2 (September 2019): multisensory integration of perceptual information.
- Kristoffer Larsen Norheim (from Aalborg University, Denmark): Doctoral stay (September-November 2019): biomechanical analysis of virtual hammering tasks.

9.5.1.1. Internships

- Sheryl Bourgaize, PhD Student, Wilfrid Laurier University, Canada (April-July 2019)
- Robyn Grunberg, Master Student, Wilfrid Laurier University, Canada (April-July 2019)

9.5.2. Visits to International Teams

9.5.2.1. Research Stays Abroad

- Simon Hilt: doctoral stay at Aalborg University Denmark (May-July 2019): biomechanical analysis of hammering tasks.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

10.1.1.1. Member of the Organizing Committees

- Marc Christie, Co-chair and co-organizer of the 8th Eurographics WICED event (Workshop on Intelligent Cinematography and Editing), 2019
- Anne-Hélène Olivier, Co-organizer of the workshop "Modeling and Animating Realistic Crowds and Humans (MARCH)" during the AIVR Conference, San Diego, USA, December 2019
- Anne-Hélène Olivier, Co-Organizer of the Symposium " Going from here to there and beyond: Fundamental theories and applications from what we have learned about human navigation of cluttered environments", ISPGR 2019, Edinburgh, Scotland, July 2019
- Anne-Hélène Olivier, Co-organizer of the workshop "Virtual Humans and Crowds in Immersive Environments" (VHCIE) during IEEE VR Conference Osaka Japan, March 2019

10.1.2. Scientific Events: Selection

10.1.2.1. Chair of Conference Program Committees

- Ludovic Hoyet, Co-Program Chair of the 19th ACM Symposium on Applied Perception, Barcelona, September 2019
- Ludovic Hoyet, Co-Workshop Chair for of the 32nd International Conference on Computer Animation and Social Agents, Paris, July 2019
- Ludovic Hoyet, Animation & Simulation Area Program Chair, International Conference on Computer Graphics Theory and Applications 2019, Prague, Czech Republic, Feb. 2019
- Marc Christie, Area Chair, ACM TVX 2019, International Conference on Interactive Experiences for Television and Online Video
- Anne-Hélène Olivier, Conference Paper co-chair, IEEE Virtual Reality Conference, Osaka Japan, March 2019

10.1.2.2. Member of the Conference Program Committees

- Ludovic Hoyet, ACM Motion Interaction in Games MIG 2019
- Marc Christie, ACM Motion Interaction in Games MIG 2019
- Marc Christie, ACM Interactive Experiences for Television and Online Video TVX 2019
- Anne-Hélène Olivier, ACM Motion Interaction in Games MIG 2019

- Anne-Hélène Olivier, ACM Symposium on Applied Perception SAP 2019
- Anne-Hélène Olivier, ACM Symposium on Virtual Reality Software and Technology VRST, Technical Papers 2019
- Franck Multon, ACM Motion Interaction in Games MIG 2019
- Franck Multon, Computer Animation and Social Agents CASA2019

10.1.2.3. Reviewer

- Ludovic Hoyet, ISVC 2019, JFIG 2019, CRDH Workshop (AIVR 2019), VHCIE Workshop (IEEE VR 2019)
- Franck Multon, CASA 2019, MIG 2019

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Anne-Hélène Olivier, Review Editor in Frontiers in Virtual Reality - Technologies for VR
- Franck Multon, Presence, MIT Press
- Franck Multon, Computer Animation and Virtual Worlds CAVW, John Wiley
- Georges Dumont, Review Editor in Frontiers in Virtual Reality - Technologies for VR

10.1.3.2. Reviewer - Reviewing Activities

- Charles Pontonnier - Applied Science, Applied ergonomics, Multibody Systems Dynamics, Sensors, Complexity
- Marc Christie, ACM SIGGRAPH 2019, ACM SIGGRAPH ASIA 2019, TVCG
- Ludovic Hoyet, IEEE Transactions on Visualization and Computer Graphics, Neuroscience, Journal of Virtual Reality and Broadcasting, Journal on Multimodal User Interfaces
- Anne-Hélène Olivier, Virtual Reality, Gait and Posture
- Franck Multon, Virtual Reality, The Visual Computer, IEEE Robotics and Automation, International Journal of Industrial Ergonomics, Journal of Sports Sciences, Journal of Ergonomics
- Armel Crétual, Clinical Biomechanics, Medical and Biological Engineering and Computing, PlosONE, Sensors
- Georges Dumont, Journal of Biomechanics

10.1.4. Invited Talks

- Marc Christie, Workshop AI in Media at European Broadcast Union, Manchester 2019.
- Anne-Hélène Olivier, Sport Unlimitech, Lyon, September 2019.

10.1.5. Scientific Expertise

- Franck Multon, expertise for I-Site NExT from Nantes for the "International Research Partnerships" call for projects
- Franck Multon, FRQNT (Canada) for the evaluation of 4 collaborative projects in Quebec on the topic "computer sciences - human-robot interaction"

10.1.6. Research Administration

- 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 is an elected member of the Inria Laboratory Board, since Sept. 2019.
- Franck Multon is the Inria national coordinator for digital sciences for Sports, since Sept 2018

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

- 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 simulation of polyarticulated systems", leader of the module, 22H, M1 Mechatronics, Ecole Normale Supérieure de Rennes, France
- Master: Charles Pontonnier, Responsible of the internships of students (L3 and M1), 15H, Ecole Normale Supérieure de Rennes, France
- Master: Charles Pontonnier, Reearch projects, 20H, Ecole Normale Supérieure de Rennes, France
- Master : Georges Dumont, Responsible of the second year of the master Engineering of complex systems, École Normale Supérieure de Rennes and Rennes 1 University, France
- Master : Georges Dumont, Mechanical simulation in Virtual reality, 36H, Master Engineering of complex systems and Mechatronics, Rennes 1 University and École Normale Supérieure de Rennes, France
- Master : Georges Dumont, Mechanics of deformable systems, 40H, Master, École Normale Supérieure de Rennes, France
- Master : Georges Dumont, oral preparation to agregation competitive exam, 20H, Master, École Normale Supérieure de Rennes, France
- Master : Georges Dumont, Vibrations in Mechanics, 10H, Master, École Normale Supérieure de Rennes, France
- Master : Georges Dumont, Finite Element method, 12H, Master, École Normale Supérieure de Rennes, France
- Master : Ludovic Hoyet, Motion Analysis and Gesture Recognition, 12h, INSA Rennes, France
- Master : Ludovic Hoyet, Réalité Virtuelle pour l'Analyse Ergonomique, Master Ingénierie et Ergonomie des Activités Physique, 21h, University Rennes 2, France
- Master : Anne-Hélène Olivier, co-leader of the APPCM Master (1 and 2) "Activités Physiques et Pathologies Chroniques et Motrices", STAPS, University Rennes2, France
- Master : Anne-Hélène Olivier, "Biostatstiques", 21H, Master 2 APPCM/IEAP, University Rennes2, France
- Master : Anne-Hélène Olivier, "Evaluation fonctionnelle des pathologies motrices", 3H 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", 10H, Master 1 IEAP/APPCM, University Rennes2, France
- Master : Anne-Hélène Olivier, "Contrôle moteur : Boucle perceptivo-motrice", 3H, Master 1IEAP, Université Rennes 2, 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, "Modelistion Animation Rendu", Master 2, 16h, leader of the module, Computer Science, University of Rennes 1, France
- Master: Marc Christie, "Advanced Computer Graphics", Master 1, 10h, leader of the module, Computer Science, ENS, France
- Licence : Franck Multon, "Ergonomie du poste de travail", Licence STAPS L2 & L3, University Rennes2, France
- Licence: Charles Pontonnier, "Lagrangian Mechanics" , leader of the module, 22H, L3 Mechatronics, Ecole Normale Supérieure de Rennes, France
- Licence: Charles Pontonnier, "Serial Robotics", leader of the module, 24H, L3 Mechatronics, , Ecole Normale Supérieure de Rennes, 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: 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

10.2.2. Supervision

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 (beginning January 2019): Nils Hareng, simulation of plausible bipedal locomotion of human and non-human primate, University Rennes2, Franck Multon, & Bruno Watier (CNRS LAAS in Toulouse)

PhD in progress (beginning January 2019): Nicolas Olivier, Adaptive Avatar Customization for Immersive Experience, Cifre InterDigital, Franck Multon, Ferran Arguelaget (Hybrid team), Quentin Avril (InterDigital), Fabien Danieau (InterDigital)

PhD in progress (beginning September 2017): Lyse Leclerc, Intérêts dans les activités physiques du rétablissement de la fonction inertielle des membres supérieurs en cas d'amputation ou d'atrophie, Armel Crétual, Diane Haering

PhD in progress (beginning September 2018) : Jean Basset, Learning Morphologically Plausible Pose Transfer, Inria, Edmond Boyer (Morpheo Inria Grenoble), Franck Multon

PhD in progress (beginning October 2018): Hiep Pham Tuan, Heterogeneous data fusion for safeguarding of cultural heritage of dance, University Reims Champagne Ardennes, Celine Loscos (University Reims), Franck Multon

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 2018): Olfa Haj Mamhoud, Monitoring de l'efficience gestuelle d'opérateurs sur postes de travail, University Rennes 2, Franck Multon, Georges Dumont, Charles Pontonnier

PhD in progress (beginning September 2019) : Claire Livet, Dynamique contrainte pour l'analyse musculo-squelettique en temps rapide : vers des méthodes alternatives d'estimation des forces musculaires mises en jeu dans le mouvement humain, Ecole normale supérieure, Georges Dumont, Charles Pontonnier

PhD in progress (beginning september 2019) : Louise Demestre, simulation MUsculo-squelettique et Structure Elastique pour le Sport (MUSES), Ecole normale supérieure, Georges Dumont, Charles Pontonnier, Nicolas Bideau, Guillaume Nicolas

PhD in progress: Diane Dewez, Avatar-Based Interaction in Virtual Reality, Oct. 2018, Ferran Argelaguet, Ludovic Hoyet, Anatole Lécuyer. Coll. with Hybrid team.

PhD in progress: Rebecca Fribourg, Enhancing Avatars in Virtual Reality through Control, Interactions and Feedback, Sept. 2017, Ferran Argelaguet, Ludovic Hoyet, Anatole Lécuyer. Coll. with Hybrid team

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 Pettré. Coll with Rainbow team

PhD in progress: Benjamin Niay, A framework for synthezing personalised human motions from motion capture data and perceptual information, Oct. 2018, Ludovic Hoyet, Anne-Hélène Olivier, Julien Pettré. Coll. with Rainbow team

PhD in progress: Alberto Jovane, Modélisation de mouvements réactifs et comportements non verbaux pour la création d'acteurs digitaux pour la réalité virtuelle, Sept. 2019, Marc Christie, Ludovic Hoyet, Claudio Pacchierotti, Julien Pettré. Coll. with Rainbow team

PhD in progress: Adèle Colas, Modélisation de comportements collectifs réactifs et expressifs pour la réalité virtuelle, Nov. 2019, Ludovic Hoyet, Anne-Hélène Olivier, Claudio Pacchierotti, Julien Pettré. Coll. with Rainbow team

PhD in progress: Hugo Brument, Toward user-adapted interactions techniques based on human locomotion laws for navigating in virtual environments, Oct. 2018, Anne-Hélène Olivier, Ferran Argelaguet, Maud Marchal. Coll. with Hybrid team

PhD in progress: Lyse Leclercq, Intérêts dans les activités physiques du rétablissement de la fonction inertielle des membres supérieurs en cas d'amputation ou d'atrophie, Sep. 2017, Armel Crétual, Diane Haering

PhD in progress: Florence Gaillard, Validation d'un protocole d'analyse 3D du mouvement des membres supérieurs en situation écologique chez les enfants ayant une Hémiplégie Cérébrale Infantile, Armel Crétual, Isabelle Bonan

PhD in progress: Karim Jamal, Effets des stimulations sensorielles par vibration des muscles du cou sur les perturbations posturales secondaires aux troubles de la représentation spatiale après un accident vasculaire cérébral, Sep 2016, Isabelle Bonan, Armel Crétual

PhD in progress: Carole Puil, Impact d'une stimulation plantaire orthétique sur la posture d'individus sains et posturalement déficients au cours de la station debout, et lors de la marche, Sep. 2018, Armel Crétual, Anne-Hélène Olivier

PhD: Charles Faure, Analyse en réalité virtuelle de la coopération lors d'une interception de balle : interaction et interférence, Sep. 2016, defense 17/01/2019, Richard Kulpa, Benoit Bideau

PhD in progress: Théo Perrin, Evaluation de la perception de l'effort en réalité virtuelle pour l'entraînement sportif, Sep. 2017, Richard Kulpa, Benoit Bideau

PhD in progress: Annabelle Limballe, Anticipation dans les sports de combat : la réalité virtuelle comme solution innovante d'entraînement, Sep. 2019, Richard Kulpa & Benoit Bideau

PhD in progress: Alexandre Vu, 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 , Sep. 2019, Richard Kulpa & Benoit Bideau

10.2.3. Juries

- PhD defense : École Nationale Supérieure d'Arts et Métiers, Samuel Hybois. « Approche numérique pour l'optimisation personnalisée des réglages d'un fauteuil roulant manuel ». 2019-10-14. Georges Dumont, President
- PhD defense : Université de Grenoble-Alpes, Chaowanan Khundam. « Évaluation de l'usage de la réalité virtuelle pour la promotion de patrimoine historique-Evaluation of the virtual reality usage for the promotion of historical heritage ». March 2019. Georges Dumont, President.
- PhD defense: University Toulouse 3, Guillaume Fumery, "Biomécanique du transport collectif de charges, vers une application clinique", january 2019. Franck Multon, Referee
- PhD defense: INSA Lyon, Samuel Carensac, "Contrôle physique de mouvement de personnages virtuels en environnement complexe", July 2019. Franck Multon, Referee
- HDR defense: Sorbonne University, Jonathan Savin, "Simulation de la variabilité du mouvement induite par la fatigue musculaire pour la conception ergonomique de postes de travail", november 2019, Franck Multon, Examiner
- HDR defense: University Lyon 1, Thomas Robert, "Analyse et simulation du maintien de l'équilibre debout chez l'humain", November 2019, Franck Multon, President

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

- Franck Multon is coordinating the national action of Inria in sports

10.3.2. Interventions

- Charles Pontonnier - "La simulation musculo-squelettique pour la robotique" - 4èmes rencontres interdisciplinaires Innovation, Robotique et Santé, December 2019
- Franck Multon - "fête de la Science" Rennes in Champs Libres for two days (October 4th and 5th) demos in the field of VR for sports for a wide public audience
- Franck Multon - "fête de la science" Paris (October 5th), stand-up talk about VR & Sports to the wide public audience, interview with "Esprit Sorcier" for a web-Tv diffusion
- Franck Multon - Sport Unlimitech Lyon, a large event about new technologies for sports (September 19-21), with keynote talks/round tables, managing the booths of Inria
- Armel Crétual - "La science sur les planches", Spectacle "Marcher!!!" avec le Collège Théophile Briant de Tinténiac et la compagnie Kali&Co
- Richard Kulpa - "La réalité virtuelle pour le sport, ludique mais encore ?", Mardis de l'Espace des Sciences, December 2019
- Richard Kulpa - Inria booth at Sport Unlimitech Lyon, a large event about new technologies for sports, September 2019
- Richard Kulpa - Booth at FIFA Women's World Cup 2019
- Anthony Sorel - Inria booth at Sport Unlimitech Lyon, a large event about new technologies for sports, September 2019

10.3.3. Internal action

- Franck Multon - "Journes scientifiques Inria", June 5-6th, plenary talk entitled "La réalité virtuelle pour le sport : un jeu vraiment sérieux ?"
- Franck Multon, Ronan Gaugne - "20 years of Immersia": Event organized by Inria Rennes/IRISA center, November 2019, plenary talks

11. Bibliography

Major publications by the team in recent years

[1] S. Cordillet, N. Bideau, B. Bideau, G. Nicolas.*Estimation of 3D Knee Joint Angles during Cycling Using Inertial Sensors Accuracy of a Novel Sensor-to-Segment Calibration Procedure Based on Pedaling Motion*, in "Sensors", 2019, vol. 19, n^{o} 11, p. 1-23 [*DOI :* 10.3390/s19112474], https://hal-univ-rennes1.archives-ouvertes.fr/hal-02160398

[2] C. Faure, A. Limballe, A. Sorel, T. Perrin, B. Bideau, R. Kulpa.*Dyadic Interference Leads to Area of Uncertainty During Face-to-Face Cooperative Interception Task*, in "Frontiers in information and communication technologies", October 2019, vol. 6, p. 1-13 [*DOI :* 10.3389/fict.2019.00020], https://hal.inria.fr/hal-02407452

[3] Q. Galvane, I.-S. Lin, F. Argelaguet Sanz, T.-Y. Li, M. Christie.*VR as a Content Creation Tool for Movie Previsualisation*, in "VR 2019 - 26th IEEE Conference on Virtual Reality and 3D User Interfaces", Osaka, France, IEEE, March 2019, p. 303-311 [*DOI :* 10.1109/VR.2019.8798181], https://hal.inria.fr/hal-02295309

[4] D. HAERING, C. PONTONNIER, N. BIDEAU, G. NICOLAS, G. DUMONT.*Using Torque-Angle and Torque-Velocity Models to Characterize Elbow Mechanical Function: Modeling and Applied Aspects*, in "Journal of Biomechanical Engineering", May 2019, vol. 141, n^{o} 8, 084501 [*DOI :* 10.1115/1.4043447], https://hal.inria.fr/hal-02074561

[5] S. D. LYNCH, A.-H. OLIVIER, B. BIDEAU, R. KULPA.*Detection of deceptive motions in rugby from visual motion cues*, in "PLoS ONE", 2019, vol. 14, n^{o} 9, e0220878 [*DOI :* 10.1371/JOURNAL.PONE.0220878], https://hal-univ-rennes1.archives-ouvertes.fr/hal-02308029

[6] A. MULLER, C. PONTONNIER, G. DUMONT.*Motion-based prediction of hands and feet contact efforts during asymmetric handling tasks*, in "IEEE Transactions on Biomedical Engineering", 2019, p. 1-11 [*DOI :* 10.1109/TBME.2019.2913308], https://hal.inria.fr/hal-02109407

[7] A. MULLER, C. PONTONNIER, P. PUCHAUD, G. DUMONT.*CusToM: a Matlab toolbox for musculoskeletal simulation*, in "Journal of Open Source Software", January 2019, vol. 4, n^{o} 33, p. 1-3 [*DOI :* 10.21105/JOSS.00927], https://hal.inria.fr/hal-01988715

[8] A. MULLER, C. PONTONNIER, X. ROBERT-LACHAINE, G. DUMONT, A. PLAMONDON.*Motion-based prediction of external forces and moments and back loading during manual material handling tasks*, in "Applied Ergonomics", 2019, p. 1-6 [*DOI :* 10.1016/J.APERGO.2019.102935], https://hal.inria.fr/hal-02268958

[9] V. RAPOS, M. CINELLI, N. SNYDER, A. CRÉTUAL, A.-H. OLIVIER.*Minimum predicted distance Applying a common metric to collision avoidance strategies between children and adult walkers*, in "Gait and Posture", 2019, vol. 72, p. 16-21 [*DOI :* 10.1016/J.GAITPOST.2019.05.016], https://hal-univ-rennes1.archives-ouvertes.fr/hal-02160399

[10] A. SOREL, P. PLANTARD, N. BIDEAU, C. PONTONNIER.*Studying fencing lunge accuracy and response time in uncertain conditions with an innovative simulator*, in "PLoS ONE", 2019, vol. 14, n^{o} 7, e0218959 [*DOI :* 10.1371/JOURNAL.PONE.0218959], https://hal.inria.fr/hal-02159746

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] S. CORDILLET.*Development of new methodologies for in situ motion analysis of cycling. Contribution of inertial measurement units for joint kinematics assessment.*, Université Rennes 2 - COMUE UNIVERSITE BRETAGNE LOIRE, December 2019, https://hal.archives-ouvertes.fr/tel-02486746

[12] C. FAURE.*Analysis in virtual reality of cooperation during ball interception: interaction and interference*, Université Rennes 2, December 2019, https://hal.archives-ouvertes.fr/tel-02430381

[13] C. PONTONNIER.*Efficient motion analysis and virtual reality methods for preventive and corrective ergonomics*, ENS Rennes, November 2019, Habilitation à diriger des recherches, https://hal.archives-ouvertes.fr/tel-02392255

Articles in International Peer-Reviewed Journal

[14] S. CORDILLET, N. BIDEAU, B. BIDEAU, G. NICOLAS.*Estimation of 3D Knee Joint Angles during Cycling Using Inertial Sensors Accuracy of a Novel Sensor-to-Segment Calibration Procedure Based on Pedaling*

Motion, in "Sensors", 2019, vol. 19, n^{o} 11, p. 1-23 [*DOI :* 10.3390/S19112474], https://hal-univ-rennes1.archives-ouvertes.fr/hal-02160398

[15] C. FAURE, A. LIMBALLE, B. BIDEAU, R. KULPA.*Virtual reality to assess and train team ball sports performance: A scoping review*, in "Journal of Sports Sciences", January 2020, vol. 38, n^{o} 2, p. 192-205 [*DOI :* 10.1080/02640414.2019.1689807], https://hal.inria.fr/hal-02442089

[16] C. FAURE, A. LIMBALLE, A. SOREL, T. PERRIN, B. BIDEAU, R. KULPA.*Dyadic Interference Leads to Area of Uncertainty During Face-to-Face Cooperative Interception Task*, in "Frontiers in information and communication technologies", October 2019, vol. 6, p. 1-13 [*DOI :* 10.3389/FICT.2019.00020], https://hal.inria.fr/hal-02407452

[17] D. HAERING, C. PONTONNIER, N. BIDEAU, G. NICOLAS, G. DUMONT.*Using Torque-Angle and Torque-Velocity Models to Characterize Elbow Mechanical Function: Modeling and Applied Aspects*, in "Journal of Biomechanical Engineering", May 2019, vol. 141, n^{o} 8, 084501 [*DOI :* 10.1115/1.4043447], https://hal.inria.fr/hal-02074561

[18] S. HYBOIS, P. PUCHAUD, M. BOURGAIN, A. LOMBART, J. BASCOU, F. LAVASTE, P. FODÉ, H. PILLET, C. SAURET.*Comparison of shoulder kinematic chain models and their influence on kinematics and kinetics in the study of manual wheelchair propulsion*, in "Medical Engineering and Physics", July 2019, vol. 69, p. 153-160 [*DOI :* 10.1016/J.MEDENGPHY.2019.06.002], https://hal.archives-ouvertes.fr/hal-02186404

[19] S. D. LYNCH, A.-H. OLIVIER, B. BIDEAU, R. KULPA.*Detection of deceptive motions in rugby from visual motion cues*, in "PLoS ONE", 2019, vol. 14, n^{o} 9, e0220878 [*DOI :* 10.1371/JOURNAL.PONE.0220878], https://hal-univ-rennes1.archives-ouvertes.fr/hal-02308029

[20] C. MARTIN, B. BIDEAU, P. TOUZARD, R. KULPA.*Identification of serve pacing strategies during five-set tennis matches*, in "International journal of Sports Science and Coaching", February 2019, vol. 14, n^{o} 1, p. 32-42 [*DOI :* 10.1177/1747954118806682], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01940810

[21] A. MULLER, C. PONTONNIER, G. DUMONT.*Motion-based prediction of hands and feet contact efforts during asymmetric handling tasks*, in "IEEE Transactions on Biomedical Engineering", 2019, p. 1-11 [*DOI :* 10.1109/TBME.2019.2913308], https://hal.inria.fr/hal-02109407

[22] A. MULLER, C. PONTONNIER, P. PUCHAUD, G. DUMONT.*CusToM: a Matlab toolbox for musculoskeletal simulation*, in "Journal of Open Source Software", January 2019, vol. 4, n^{o} 33, p. 1-3 [*DOI :* 10.21105/JOSS.00927], https://hal.inria.fr/hal-01988715

[23] A. MULLER, C. PONTONNIER, X. ROBERT-LACHAINE, G. DUMONT, A. PLAMONDON.*Motion-based prediction of external forces and moments and back loading during manual material handling tasks*, in "Applied Ergonomics", 2019, p. 1-6 [*DOI :* 10.1016/J.APERGO.2019.102935], https://hal.inria.fr/hal-02268958

[24] C. PONTONNIER, C. LIVET, A. MULLER, A. SOREL, G. DUMONT, N. BIDEAU.*Ground Reaction Forces and Moments Prediction of Challenging Motions: Fencing Lunges*, in "Computer Methods in Biomechanics and Biomedical Engineering", 2019, https://hal.archives-ouvertes.fr/hal-02142288

[25] P. PUCHAUD, G. DUMONT, N. BIDEAU, C. PONTONNIER.*Knee Torque Generation Capacities Modelled With Physiological Torque-Angle-Velocity Relationships*, in "Computer Methods in Biomechanics and Biomedical Engineering", 2019, p. 1-2, https://hal.inria.fr/hal-02162221

[26] V. RAPOS, M. CINELLI, N. SNYDER, A. CRÉTUAL, A.-H. OLIVIER.*Minimum predicted distance Applying a common metric to collision avoidance strategies between children and adult walkers*, in "Gait and Posture", 2019, vol. 72, p. 16-21 [*DOI :* 10.1016/J.GAITPOST.2019.05.016], https://hal-univ-rennes1.archives-ouvertes.fr/hal-02160399

[27] A. SOREL, P. PLANTARD, N. BIDEAU, C. PONTONNIER.*Studying fencing lunge accuracy and response time in uncertain conditions with an innovative simulator*, in "PLoS ONE", 2019, vol. 14, n^{o} 7, e0218959 [*DOI :* 10.1371/JOURNAL.PONE.0218959], https://hal.inria.fr/hal-02159746

[28] P. TOUZARD, R. KULPA, B. BIDEAU, B. MONTALVAN, C. MARTIN.*Biomechanical analysis of the "waiter's serve" on upper limb loads in young elite tennis players*, in "European Journal of Sport Science", July 2019, vol. 19, n^{o} 6, p. 765-773 [*DOI :* 10.1080/17461391.2018.1539527], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01940362

International Conferences with Proceedings

[29] J. BASSET, S. WUHRER, E. BOYER, F. MULTON.*Contact Preserving Shape Transfer For Rigging-Free Motion Retargeting*, in "MIG 2019 - ACM SIGGRAPH Conference Motion Interaction and Games", Newcastle Upon Tyne, United Kingdom, ACM, October 2019, p. 1-10 [*DOI :* 10.1145/3359566.3360075], https://hal.archives-ouvertes.fr/hal-02293308

[30] F. BERTON, A.-H. OLIVIER, J. BRUNEAU, L. HOYET, J. PETTRÉ.*Studying Gaze Behaviour During Collision Avoidance With a Virtual Walker: Influence of the Virtual Reality Setup*, in "VR 2019 - 26th IEEE Conference on Virtual Reality and 3D User Interfaces", Osaka, Japan, IEEE, March 2019, p. 717-725 [*DOI :* 10.1109/VR.2019.8798204], https://hal.inria.fr/hal-02058360

[31] A. BRUCKERT, Y. H. LAM, M. CHRISTIE, O. LE MEUR.*Deep learning for inter-observer congruency prediction*, in "ICIP 2019 - IEEE International Conference on Image Processing", Taipei, Taiwan, IEEE, September 2019, p. 3766-3770 [*DOI :* 10.1109/ICIP.2019.8803596], https://hal.inria.fr/hal-02333013

[32] H. BRUMENT, I. PODKOSOVA, H. KAUFMANN, A.-H. OLIVIER, F. ARGELAGUET SANZ.*Virtual vs. Physical Navigation in VR: Study of Gaze and Body Segments Temporal Reorientation Behaviour*, in "VR 2019 - 26th IEEE Conference on Virtual Reality and 3D User Interfaces", Osaka, Japan, IEEE, March 2019, p. 680-689 [*DOI :* 10.1109/VR.2019.8797721], https://hal.inria.fr/hal-02058365

[33] J. G. DA SILVA FILHO, A.-H. OLIVIER, A. CRÉTUAL, J. PETTRÉ, T. FRAICHARD.*Effective Human-Robot Collaboration in near symmetry collision scenarios*, in "RO-MAN 2019 - 28th IEEE International Conference on Robot & Human Interactive Communication", New Dehli, India, IEEE, October 2019, p. 1-8, https://hal.inria.fr/hal-02267705

[34] C. DEROO, A. MONTUWY, B. DEGRAEVE, J.-M. AUBERLET, M.-A. GRANIÉ, A.-H. OLIVIER.*Evitement de collision entre deux piétons sur un trottoir étroit*, in "SOFPEL 2019 - 26ème congrès de la Société Francophone Posture, Équilibre et Locomotion", Montréal, Canada, December 2019, https://hal.inria.fr/hal-02373619

[35] D. DEWEZ, R. FRIBOURG, F. ARGELAGUET SANZ, L. HOYET, D. MESTRE, M. SLATER, A. LÉCUYER.*Influence of Personality Traits and Body Awareness on the Sense of Embodiment in Virtual Reality*, in "ISMAR 2019 - 18th IEEE International Symposium on Mixed and Augmented Reality", Beijin, China, October 2019, p. 1-12, https://hal.inria.fr/hal-02385783

[36] Q. GALVANE, I.-S. LIN, F. ARGELAGUET SANZ, T.-Y. LI, M. CHRISTIE.*VR as a Content Creation Tool for Movie Previsualisation*, in "VR 2019 - 26th IEEE Conference on Virtual Reality and 3D User Interfaces", Osaka, France, IEEE, March 2019, p. 303-311 [*DOI :* 10.1109/VR.2019.8798181], https://hal.inria.fr/hal-02295309

[37] L. HOYET, P. PLANTARD, A. SOREL, R. KULPA, F. MULTON.*Influence of Motion Speed on the Perception of Latency in Avatar Control*, in "AIVR 2019 - 2nd IEEE International Conference on Artificial Intelligence & Virtual Reality", San-Diego, United States, December 2019, https://hal.inria.fr/hal-02335326

[38] B. NIAY, A.-H. OLIVIER, J. PETTRÉ, L. HOYET.*The Influence of Step Length to Step Frequency Ratio on the Perception of Virtual Walking Motions*, in "MIG 2019 - 12th annual ACM SIGGRAPH conference on Motion, Interaction and Games", Newcastle upon Tyne, United Kingdom, ACM Press, October 2019, p. 1-2 [*DOI :* 10.1145/3359566.3364687], https://hal.archives-ouvertes.fr/hal-02378300

[39] A.-H. OLIVIER, N. LE BORGNE, M. BABEL, A. CRÉTUAL, J. PETTRÉ.*Evitement de collision entre un piéton et une personne sur un fauteuil roulant motorisé*, in "SOFPEL 2019 - 26ème congrès de la Société Francophone Posture, Équilibre et Locomotion", Montréal, Canada, December 2019, https://hal.inria.fr/hal-02373525

[40] P. PUCHAUD, G. DUMONT, N. BIDEAU, C. PONTONNIER.*A case study with custom : a comparison of normal and altered gait with an ankle brace*, in "ESB 2019 - 25th Congress of the European Society of Biomechanics", Vienne, Austria, July 2019, 1, https://hal.inria.fr/hal-02088913

[41] C. PUIL, A. CRÉTUAL, A.-H. OLIVIER.*Impact of a thin plantar orthopaedic insert on posture and locomotion*, in "ENPODHE 2019 - Conference European Network of Podiatry in Higher Education", Bruxelles, Belgium, March 2019, https://hal.inria.fr/hal-02396515

National Conferences with Proceeding

[42] T. DUVERNE, T. ROUGNANT, F. LE YONDRE, F. BERTON, J. BRUNEAU, L. HOYET, J. PETTRÉ, A.-H. OLIVIER.*Analyse des réactions corporelles à la transgression des normes de proxémie en contexte de spectacle sportif : expérimentations en situations réelles et virtuelles*, in "ACAPS 2019 - 18ème congrès de l'Association des Chercheurs en Activités Physiques et Sportives", Paris, France, October 2019, https://hal.inria.fr/hal-02393360

Conferences without Proceedings

[43] S. CORDILLET, N. BIDEAU, V. MADOUAS, M. MÉNARD, B. BIDEAU, G. NICOLAS.*Lower limb muscle fatigue during 4-km track cycling time trial in high level women cyclist*, in "24th Annual Congress of the European College of Sport Science ECSS", Prague, Czech Republic, July 2019, https://hal.archives-ouvertes.fr/hal-02429231

[44] S. LYNCH, R. KULPA, L. A. MEERHOFF, A. SOREL, J. PETTRÉ, A.-H. OLIVIER.*Collision avoidance between walkers with a twist: strategies for curvilinear and rectilinear paths*, in "ISPGR 2019 - Conference of

the International Society for Posture & Gait Research", Edinburgh, United Kingdom, June 2019, https://hal.inria.fr/hal-02058340

[45] S. LYNCH, A.-H. OLIVIER, B. BIDEAU, R. KULPA.*Global motion visualisation for detection of deceptive motion in rugby*, in "24th Annual Congress of the European College of Sport Science ECSS", Prague, Czech Republic, July 2019, https://hal.inria.fr/hal-02421389

[46] B. NIAY, A.-H. OLIVIER, J. PETTRÉ, L. HOYET.*The Influence of Step Length to Step Frequency Ratio on the Perception of Virtual Walking Motions*, in "ACM SIGGRAPH Symposium on Applied Perception", Barcelone, Spain, September 2019, https://hal.archives-ouvertes.fr/hal-02395303

[47] A.-H. OLIVIER, N. LE BORGNE, M. BABEL, A. CRÉTUAL, J. PETTRÉ.*Collision avoidance between a walker and a person on an electric powered wheelchair*, in "ISPGR 2019 - Conference of the International Society for Posture & Gait Research", Edinburgh, United Kingdom, June 2019, https://hal.inria.fr/hal-02058342

Scientific Books (or Scientific Book chapters)

[48] P. PLANTARD, H. P. SHUM, F. MULTON.*Motion analysis of work conditions using commercial depth cameras in real industrial conditions*, in "DHM and Posturography", Elsevier, 2019, p. 673-682 [*DOI :* 10.1016/B978-0-12-816713-7.00052-0], https://hal.inria.fr/hal-02367500

Other Publications

[49] A. BRUCKERT, H. R. TAVAKOLI, Z. LIU, M. CHRISTIE, O. LE MEUR.*Deep Saliency Models : the Quest for the Loss Function*, August 2019, working paper or preprint, https://hal.inria.fr/hal-02264898

[50] C. DEROO, A. MONTUWY, B. DEGRAEVE, J.-M. AUBERLET, A.-H. OLIVIER, M.-A. GRANIÉ.*Pedestrian collision avoidance on narrow sidewalk:a meeting between psychology and virtual reality*, January 2019, TRB 2019 - Annual Meeting Transportation Research Board, Poster, https://hal-univ-rennes1.archives-ouvertes.fr/hal-02396553

[51] R. GRUNDBERG, M. CINELLI, S. BOURGAIZE, A. CRÉTUAL, A.-H. OLIVIER.*Collision avoidance strategies in older adults*, December 2019, SOFPEL 2019 - 26ème congrès de la Société Francophone Posture, Équilibre et Locomotion, Poster, https://hal.inria.fr/hal-02373665

[52] C. PUIL, A. CRÉTUAL, A.-H. OLIVIER.*Impact of a thin plantar orthopaedic insert on posture and locomotion*, December 2019, SOFPEL 2019 - 26ème congrès de la Société Francophone Posture, Équilibre et Locomotion, Poster, https://hal.inria.fr/hal-02373802

[53] C. PUIL, A.-H. OLIVIER, A. CRÉTUAL.*Impact of a thin plantar orthopaedic insert on posture and locomotion*, June 2019, ISPGR 2019 - Conference of the International Society for Posture & Gait Research, Poster, https://hal.inria.fr/hal-02058347

[54] V. RAPOS, M. CINELLI, N. SNYDER, A. CRÉTUAL, A.-H. OLIVIER.*Minimum Predicted Distance: applying a common metric to collision avoidance strategies between typically developing children and adult walkers*, June 2019, ISPGR 2019 - Conference of the International Society for Posture & Gait Research, Poster, https://hal.inria.fr/hal-02058358

[55] N. SNYDER, M. CINELLI, V. RAPOS, A. CRÉTUAL, A.-H. OLIVIER.*Collision avoidance between two walkers: Reduced avoidance behaviour in previously concussed athletes*, June 2019, ISPGR 2019 - Conference of the International Society for Posture & Gait Research (ISPGR), Poster, https://hal.inria.fr/hal-02058359

Project-Team MINGUS

MultI-scale Numerical Geometric Schemes

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

Project-Team MINGUS

Creation of the Team: 2018 January 01, updated into Project-Team: 2018 August 01

Keywords:

Computer Science and Digital Science:

A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.2. - Stochastic Modeling
A6.1.4. - Multiscale modeling
A6.2.1. - Numerical analysis of PDE and ODE
A6.2.7. - High performance computing

Other Research Topics and Application Domains:

B4.2.2. - Fusion
B5.11. - Quantum systems
B9.5.2. - Mathematics

1. Team, Visitors, External Collaborators

Research Scientists

Nicolas Crouseilles [Team leader, Inria, Senior Researcher, HDR]
Philippe Chartier [Inria, Senior Researcher, HDR]
Erwan Faou [Inria, Senior Researcher, HDR]
Mohammed Lemou [CNRS, Senior Researcher, HDR]

Faculty Members

François Castella [Univ de Rennes I, Professor, HDR]
Arnaud Debussche [Ecole normale supérieure de Rennes and IUF, Professor, HDR]
Florian Méhats [Univ de Rennes I, Professor, HDR]

External Collaborator

Pierre Navaro [CNRS]

Technical Staff

Yves Mocquard [Inria, Engineer, from Oct 2019]

PhD Students

Gregoire Barrué [Ecole normale supérieure de Rennes, PhD Student, from Sep 2019]
Joackim Bernier [Univ de Rennes I, PhD Student, until Aug 2019]
Quentin Chauleur [Ecole normale supérieure de Rennes, PhD Student, from Sep 2019]
Josselin Massot [Univ de Rennes I, PhD Student]
Angelo Rosello [Ecole normale supérieure de Rennes, PhD Student]
Leopold Tremant [Inria, PhD Student]

Post-Doctoral Fellows

Guillaume Morel [Inria, Post-Doctoral Fellow, until Sep 2019]
Xiaofei Zhao [Inria, Post-Doctoral Fellow, until Jan 2019]

Visiting Scientists

Fernando Casas [University Jaume I, Spain, until Jan 2019]
Yingzhe Li [Chinese Academy of Sciences, China, from Feb 2019]

Administrative Assistants

Marie-Noëlle Georgeault [Inria, Administrative Assistant, from Aug 2019]

Stéphanie Lemaile [Inria, Administrative Assistant, until Jul 2019]

2. Overall Objectives

2.1. Presentation

In applications involving complex physics, such as plasmas and nanotechnologies, numerical simulations serve as a prediction tool supplementing real experiments and are largely endorsed by engineers or researchers. Their performances rely not only on computational power, but also on the efficiency of the underlying numerical method and the complexity of the underlying models. The contribution of applied mathematics is then required, on the one hand for a better understanding of qualitative properties and a better identification of the different regimes present in the model, and on the other hand, for a more sounded construction of new models based on asymptotic analysis. This mathematical analysis is expected to greatly impact the design of *multiscale* numerical schemes.

The proposed research group MINGuS will be dedicated to the mathematical and numerical analysis of (possibly stochastic) partial differential equations (PDEs), originating from plasma physics and nanotechnologies, with emphasis on *multiscale* phenomena either of **highly-oscillatory**, of **dissipative** or **stochastic** types. These equations can be also encountered in applications to rarefied gas dynamics, radiative transfer, population dynamics or laser propagation, for which the multiscale character is modelled by a scale physical parameter ε.

Producing accurate solutions of multiscale equations is extremely challenging owing to severe restrictions to the numerical methods imposed by fast (or stiff) dynamics. *Ad-hoc* numerical methods should aim at capturing the slow dynamics solely, instead of resolving finely the stiff dynamics at a formidable computational cost. At the other end of the spectrum, the separation of scales -as required for numerical efficiency- is envisaged in asymptotic techniques, whose purpose is to describe the model in the limit where the small parameter ε tends to zero. MINGuS aspires to accommodate sophisticated tools of mathematical analysis and heuristic numerical methods in order to produce simultaneously rich asymptotic models and efficient numerical methods.
To be more specific, MINGuS aims at finding, implementing and analysing new multiscale numerical schemes for the following physically relevant multiscale problems:

- **Highly-oscillatory Schrödinger equation for nanoscale physics:** In quantum mechanics, the Schrödinger equation describes how the quantum state of some physical system changes with time. Its mathematical and numerical study is of paramount importance to fundamental and applied physics in general. We wish to specifically contribute to the mathematical modeling and the numerical simulation of confined quantum mechanical systems (in one or more space dimensions) possibly involving stochastic terms. Such systems are involved in quantum semi-conductors or atom-chips, as well as in cold atom physics (Bose-Einstein condensates) or laser propagation in optical fibers.

  The prototypical equation is written

$$\boxed{i\varepsilon\partial_t\psi^\varepsilon = \frac{\varepsilon^2}{\beta}\Delta\psi^\varepsilon + |\psi^\varepsilon|^2\psi^\varepsilon + \psi^\varepsilon\xi} \tag{40}$$

where the function $\psi^\varepsilon = \psi^\varepsilon(t,x) \in \mathbb{C}$ depends on time $t \geq 0$ and position $x \in \mathbb{R}^3$, $\xi = \xi(x,t)$ is a white noise and where the small parameter ε is the Planck's constant describing the microscopic/macroscopic ratio. The limit $\varepsilon \to 0$ is referred to as the semi-classical limit. The regime $\varepsilon = 1$ and $\beta \to 0$ (this can be for instance the relative length of the optical fiber) is highly-oscillatory. The noise ξ acts as a potential, it may represent several external perturbations. For instance temperature effects in Bose-Einstein condensation or amplification in optical fibers. The highly oscillatory regime combined with noise introduces new challenges in the design of efficient schemes.

- **Highly-oscillatory or highly-dissipative kinetic equations:** Plasma is sometimes considered as the fourth state of matter, obtained for example by bringing a gas to a very high temperature. A globally neutral gas of neutral and charged particles, called plasma, is then obtained and is described by a kinetic equation as soon as collective effects dominate as compared to binary collisions. A situation of major importance is magnetic fusion in which collisions are not predominant. In order to confine such a plasma in devices like tokamaks (ITER project) or stellarators, a large magnetic field is used to endow the charged particles with a cyclotronic motion around field lines. Note that kinetic models are also widely used for modeling plasmas in earth magnetosphere or in rarefied gas dynamics.

  Denoting $f^\varepsilon = f^\varepsilon(t,x,v) \in \mathbb{R}^+$ the distribution function of charged particles at time $t \geq 0$, position $x \in \mathbb{R}^3$ and velocity $v \in \mathbb{R}^3$, a typical kinetic equation for f^ε reads

$$\boxed{\partial_t f^\varepsilon + v \cdot \nabla_x f^\varepsilon + \big(E + \tfrac{1}{\varepsilon}(v \times B)\big) \cdot \nabla_v f^\varepsilon = \tfrac{1}{\beta} Q(f^\varepsilon) + f^\varepsilon m^\varepsilon} \tag{41}$$

  where (E,B) is the electro-magnetic field (which may itself depend on f through Maxwell's equations), m^ε is a random process (which may describe absorption or creation of particles) and Q is a collision operator. The dimensionless parameters ε, β are related to the cyclotronic frequency and the mean free path. Limits $\varepsilon \to 0$ and $\beta \to 0$ do not share the same character (the former is oscillatory and the latter is dissipative) and lead respectively to gyrokinetic and hydrodynamic models. The noise term m^ε is correlated in space and time. At the limit $\varepsilon \to 0$, it converges formally to a white noise and stochastic PDEs are obtained.

MINGUS project is the follow-up of IPSO, ending in december in 2017. IPSO original aim was to extend the analysis of geometric schemes from ODEs to PDEs. During the last evaluation period, IPSO also considered the numerical analysis of geometric schemes for (S)PDEs, possibly including multiscale phenomena. Breakthrough results [36], [38], [39], [42] have been recently obtained which deserve to be deepened and extended. It thus appears quite natural to build the MINGUS team upon these foundations.

The objective of MINGUS is twofold: the construction and the analysis of numerical schemes (such as “Uniformly Accurate numerical schemes", introduced by members of the IPSO project) for multiscale (S)PDEs originating from physics. In turn, this requires (i) a deep mathematical understanding of the (S)PDEs under consideration and (ii) a strong involvement into increasingly realistic problems, possibly resorting to parallel computing. For this aspect, we intend to benefit from the Inria Selalib software library which turns out to be the ideal complement of our activities.

3. Research Program

3.1. Research Program

The MINGUS project is devoted to the mathematical and numerical analysis of models arising in plasma physics and nanotechnology. The main goal is to construct and analyze numerical methods for the approximation of PDEs containing multiscale phenomena. Specific multiscale numerical schemes will be proposed and analyzed in different regimes (namely highly-oscillatory and dissipative). The ultimate goal is to dissociate the physical parameters (generically denoted by ε) from the numerical parameters (generically denoted by h) with a uniform accuracy. Such a task requires mathematical prerequisite of the PDEs.

Then, for a given stiff (highly-oscillatory or dissipative) PDE, the methodology of the MINGUS team will be the following

- Mathematical study of the asymptotic behavior of multiscale models.
  This part involves averaging and asymptotic analysis theory to derive asymptotic models, but also long-time behavior of the considered models.
- Construction and analysis of multiscale numerical schemes.
  This part is the core of the project and will be deeply inspired for the mathematical prerequisite. In particular, our ultimate goal is the design of *Uniformly Accurate* (UA) schemes, whose accuracy is independent of ε.
- Validation on physically relevant problems.
  The last goal of the MINGUS project is to validate the new numerical methods, not only on toy problems, but also on realistic models arising in physics of plasmas and nanotechnologies. We will benefit from the Selalib software library which will help us to scale-up our new numerical methods to complex physics.

3.1.1. Dissipative problems

In the dissipative context, the asymptotic analysis is quite well understood in the deterministic case and multiscale numerical methods have been developed in the last decades. Indeed, the so-called Asymptotic-Preserving schemes has retained a lot of attention all over the world, in particular in the context of collisional kinetic equations. But, there is still a lot of work to do if one is interesting in the derivation high order asymptotic models, which enable to capture the original solution for all time. Moreover, this analysis is still misunderstood when more complex systems are considered, involving non homogeneous relaxation rates or stochastic terms for instance. Following the methodology we aim at using, we first address the mathematical analysis before deriving multiscale efficient numerical methods.

A simple model of dissipative systems is governed by the following differential equations

$$\left\{\begin{array}{lcll} \dfrac{dx^\varepsilon(t)}{dt} & = & \mathcal{G}(x^\varepsilon(t), y^\varepsilon(t)), & x^\varepsilon(0) = x_0, \\ \\ \dfrac{dy^\varepsilon(t)}{dt} & = & -\dfrac{y^\varepsilon(t)}{\varepsilon} + \mathcal{H}(x^\varepsilon(t), y^\varepsilon(t)), & y^\varepsilon(0) = y_0, \end{array}\right. \tag{42}$$

for given initial condition $(x_0, y_0) \in \mathbb{R}^2$ and given smooth functions $\mathcal{G}, \mathcal{H}$ which possibly involve stochastic terms.

3.1.1.1. *Asymptotic analysis of dissipative PDEs (F. Castella, P. Chartier, A. Debussche, E. Faou, M. Lemou)*

Derivation of asymptotic problems
Our main goal is to analyze the asymptotic behavior of dissipative systems of the form (3) when ε goes to zero. The *center manifold theorem* [35] is of great interest but is largely unsatisfactory from the following points of view

- a constructive approach of h and x_0^ε is clearly important to identify the high-order asymptotic models: this would require expansions of the solution by means of B-series or word-series [37] allowing the derivation of error estimates between the original solution and the asymptotic one.
- a better approximation of the transient phase is strongly required to capture the solution for small time: extending the tools developed in averaging theory, the main goal is to construct a suitable change of variable which enables to approximate the original solution for all time.

Obviously, even at the ODE level, a deep mathematical analysis has to be performed to understand the asymptotic behavior of the solution of (3). But, the same questions arise at the PDE level. Indeed, one certainly expects that dissipative terms occurring in collisional kinetic equations (2) may be treated theoretically along this perspective. The key new point indeed is to see the center manifold theorem as a change of variable in the space on unknowns, while the standard point of view leads to considering the center manifold as an asymptotic object.

Stochastic PDEs
We aim at analyzing the asymptotic behavior of stochastic collisional kinetic problems, that is equation of the type (2). The noise can describe creation or absorption (as in (2)), but it may also be a forcing term or a random magnetic field. In the parabolic scaling, one expect to obtain parabolic SPDEs at the limit. More precisely, we want to understand the fluid limits of kinetic equations in the presence of noise. The noise is smooth and non delta correlated. It contains also a small parameter and after rescaling converges formally to white noise. Thus, this adds another scale in the multiscale analysis. Following the pioneering work [38], some substantial progresses have been done in this topic.

More realistic problems may be addressed such as high field limit describing sprays, or even hydrodynamic limit. The full Boltzmann equation is a very long term project and we wish to address simpler problems such as convergences of BGK models to a stochastic Stokes equation.

The main difficulty is that when the noise acts as a forcing term, which is a physically relevant situation, the equilibria are affected by the noise and we face difficulties similar to that of high field limit problems. Also, a good theory of averaging lemma in the presence of noise is lacking. The methods we use are generalization of the perturbed test function method to the infinite dimensional setting. We work at the level of the generator of the infinite dimensional process and prove convergence in the sense of the martingale problems. A further step is to analyse the speed of convergence. This is a prerequisite if one wants to design efficient schemes. This requires more refined tools and a good understanding of the Kolmogorov equation.

3.1.1.2. Numerical schemes for dissipative problems (All members)

The design of numerical schemes able to reproduce the transition from the microscopic to macroscopic scales largely matured with the emergence of the Asymptotic Preserving schemes which have been developed initially for collisional kinetic equations (actually, for solving (2) when $\beta \to 0$). Several techniques have flourished in the last decades. As said before, AP schemes entail limitations which we aim at overcoming by deriving

- AP numerical schemes whose numerical cost diminishes as $\beta \to 0$,
- Uniformly accurate numerical schemes, whose accuracy is independent of β.

Time diminishing methods
The main goal consists in merging Monte-Carlo techniques [33] with AP methods for handling *automatically* multiscale phenomena. As a result, we expect that the cost of the so-obtained method decreases when the asymptotic regime is approached; indeed, in the collisional (i.e. dissipative) regime, the deviational part becomes negligible so that a very few number of particles will be generated to sample it. A work in this direction has been done by members of the team.

We propose to build up a method which permits to realize the transition from the microscopic to the macroscopic description without domain decomposition strategies which normally oblige to fix and tune an interface in the physical space and some threshold parameters. Since it will permit to go over domain decomposition and AP techniques, this approach is a very promising research direction in the numerical approximation of multiscale kinetic problems arising in physics and engineering.

Uniformly accurate methods

To overcome the accuracy reduction observed in AP schemes for intermediate regimes, we intend to construct and analyse multiscale numerical schemes for (3) whose error is uniform with respect to ε. The construction of such a scheme requires a deep mathematical analysis as described above. Ideally one would like to develop schemes that preserve the center manifold (without computing the latter!) as well as schemes that resolve numerically the stiffness induced by the fast convergence to equilibrium (the so-called transient phase). First, our goal is to extend the strategy inspired by the central manifold theorem in the ODE case to the PDE context, in particular for collisional kinetic equations (2) when $\beta \to 0$. The design of Uniformly Accurate numerical schemes in this context would require to generalize two-scale techniques introduced in the framework of highly-oscillatory problems [36].

Multiscale numerical methods for stochastic PDEs
AP schemes have been developed recently for kinetic equations with noise in the context of Uncertainty Quantification UQ [41]. These two aspects (multiscale and UQ) are two domains which usually come within the competency of separate communities. UQ has drawn a lot of attention recently to control the propagation of data pollution; undoubtedly UQ has a lot of applications and one of our goals will be to study how sources of uncertainty are amplified or not by the multiscale character of the model. We also wish to go much further and by developing AP schemes when the noise is also rescaled and the limit is a white noise driven SPDE, as described in section (3.1.1.1). For simple nonlinear problem, this should not present much difficulties but new ideas will definitely be necessary for more complicated problems when noise deeply changes the asymptotic equation.

3.1.2. Highly-oscillatory problems

As a generic model for highly-oscillatory systems, we will consider the equation

$$\frac{du^\varepsilon(t)}{dt} = \mathcal{F}(t/\varepsilon, u^\varepsilon(t)), \quad u^\varepsilon(0) = u_0, \tag{43}$$

for a given u_0 and a given periodic function $\mathcal{F}$ (of period P w.r.t. its first variable) which possibly involves stochastic terms. Solution u^ε exhibits high-oscillations in time superimposed to a slow dynamics. Asymptotic techniques -resorting in the present context to *averaging* theory [45]- allow to decompose

$$u^\varepsilon(t) = \Phi_{t/\varepsilon} \circ \Psi_t \circ \Phi_0^{-1}(u_0), \tag{44}$$

into a fast solution component, the εP-periodic change of variable $\Phi_{t/\varepsilon}$, and a slow component, the flow Ψ_t of a non-stiff *averaged* differential equation. Although equation (5) can be satisfied only up to a small remainder, various methods have been recently introduced in situations where (4) is posed in $\mathbb{R}^n$ or for the Schrödinger equation (1).

In the asymptotic behavior $\varepsilon \to 0$, it can be advantageous to replace the original singularly perturbed model (for instance (1) or (2)) by an approximate model which does not contain stiffness any longer. Such reduced models can be derived using asymptotic analysis, namely averaging methods in the case of highly-oscillatory problems. In this project, we also plan to go beyond the mere derivation of limit models, by searching for better approximations of the original problem. This step is of mathematical interest *per se* but it also paves the way of the construction of multiscale numerical methods.

3.1.2.1. Asymptotic analysis of highly-oscillatory PDEs (All members)

Derivation of asymptotic problems
We intend to study the asymptotic behavior of highly-oscillatory evolution equations of the form (4) posed in an infinite dimensional Banach space.

Recently, the stroboscopic averaging has been extended to the PDE context, considering nonlinear Schrödinger equation (1) in the highly-oscillatory regime. A very exciting way would be to use this averaging strategy for highly-oscillatory kinetic problem (2) as those encountered in strongly magnetized plasmas. This turns out to be a very promising way to re-derive gyrokinetic models which are the basis of tokamak simulations in the physicists community. In contract with models derived in the literature (see [34]) which only capture the average with respect to the oscillations, this strategy allows for the complete recovery of the exact solution from the asymptotic (non stiff) model. This can be done by solving companion transport equation that stems naturally from the decomposition (5).

Long-time behavior of Hamiltonian systems

The study of long-time behavior of nonlinear Hamiltonian systems have received a lot of interest during the last decades. It enables to put in light some characteristic phenomena in complex situations, which are beyond the reach of numerical simulations. This kind of analysis is of great interest since it can provide very precise properties of the solution. In particular, we will focus on the dynamics of nonlinear PDEs when the initial condition is close to a stationary solution. Then, the long-time behavior of the solution is studied through mainly three axis

- *linear stability*: considering the linearized PDE, do we have stability of a stationary solution ? Do we have linear Landau damping around stable non homogeneous stationary states?
- *nonlinear stability*: under a criteria, do we have stability of a stationary solution in energy norm like in [42], and does this stability persist under numerical discretization? For example one of our goals is to address the question of the existence and stability of discrete travelling wave in space and time.
- do we have existence of damped solutions for the full nonlinear problem ? Around homogeneous stationary states, solutions scatter towards a modified stationary state (see [43], [39]). The question of existence of Landau damping effects around non homogeneous states is still open and is one of our main goal in the next future.

Asymptotic behavior of stochastic PDEs

The study of SPDEs has known a growing interest recently, in particular with the fields medal of M. Hairer in 2014. In many applications such as radiative transfer, molecular dynamics or simulation of optical fibers, part of the physical interactions are naturally modeled by adding supplementary random terms (the noise) to the initial deterministic equations. From the mathematical point of view, such terms change drastically the behavior of the system.

- In the presence of noise, highly-oscillatory dispersive equations presents new problems. In particular, to study stochastic averaging of the solution, the analysis of the long time behavior of stochastic dispersive equations is required, which is known to be a difficult problem in the general case. In some cases (for instance highly-oscillatory Schrödinger equation (1) with a time white noise in the regime $\varepsilon \ll 1$), it is however possible to perform the analysis and to obtain averaged stochastic equations. We plan to go further by considering more difficult problems, such as the convergence of a stochastic Klein-Gordon-Zakharov system to as stochastic nonlinear Schrödinger equation.
- The long-time behavior of stochastic Schrödinger equations is of great interest to analyze mathematically the validity of the Zakharov theory for wave turbulence (see [44]). The problem of wave turbulence can be viewed either as a deterministic Hamiltonian PDE with random initial data or a randomly forced PDEs where the stochastic forcing is concentrated in some part of the spectrum (in this sense it is expected to be a hypoelliptic problem). One of our goals is to test the validity the Zakharov equation, or at least to make rigorous the spectrum distribution spreading observed in the numerical experiments.

3.1.2.2. Numerical schemes for highly-oscillatory problems (All members)

This section proposes to explore numerical issues raised by highly-oscillatory nonlinear PDEs for which (4) is a prototype. Simulating a highly-oscillatory phenomenon usually requires to adapt the numerical parameters in order to solve the period of size ε so as to accurately simulate the solution over each period, resulting in a unacceptable execution cost. Then, it is highly desirable to derive numerical schemes able to advance the solution by a time step independent of ε. To do so, our goal is to construct *Uniformly Accurate* (UA) numerical schemes, for which the numerical error can be estimated by Ch^p (h being any numerical parameters) with C independent of ε and p the order of the numerical scheme.

Recently, such numerical methods have been proposed by members of the team in the highly-oscillatory context [36]. They are mainly based on a separation of the fast and slow variables, as suggested by the decomposition (5). An additional ingredient to prove the uniformly accuracy of the method for (4) relies on the search for an appropriate initial data which enables to make the problem smooth with respect to ε.

Such an approach is assuredly powerful since it provides a numerical method which enables to capture the high oscillations in time of the solution (and not only its average) even with a large time step. Moreover, in the asymptotic regime, the potential gain is of order $1/\varepsilon$ in comparison with standard methods, and finally averaged models are not able to capture the intermediate regime since they miss important information of the original problem. We are strongly convinced that this strategy should be further studied and extended to cope with some other problems. The ultimate goal being to construct a scheme for the original equation which degenerates automatically into a consistent approximation of the averaged model, without resolving it, the latter can be very difficult to solve.

- Space oscillations:
  When rapidly oscillating coefficients in **space** (*i.e.* terms of the form $a(x, x/\varepsilon)$) occur in elliptic or parabolic equations, homogenization theory and numerical homogenization are usually employed to handle the stiffness. However, these strategies are in general not accurate for all $\varepsilon \in]0, 1]$. Then, the construction of numerical schemes which are able to handle both regimes in an uniform way is of great interest. Separating fast and slow *spatial* scales merits to be explored in this context. The delicate issue is then to extend the choice suitable initial condition to an *appropriate choice of boundary conditions* of the augmented problem.
- Space-time oscillations:
  For more complex problems however, the recent proposed approaches fail since the main oscillations cannot be identified explicitly. This is the case for instance when the magnetic field B depends on t or x in (2) but also for many other physical problems. We then have to deal with the delicate issue of space-time oscillations, which is known to be a very difficult problem from a mathematical and a numerical point of view. To take into account the space-time mixing, a periodic motion has to be detected together with a phase S which possibly depends on the time and space variables. These techniques originate from **geometric optics** which is a very popular technique to handle highly-frequency waves.
- Geometrical properties:
  The questions related to the geometric aspects of multiscale numerical schemes are of crucial importance, in particular when long-time simulations are addressed (see [40]). Indeed, one of the main questions of geometric integration is whether intrinsic properties of the solution may be passed onto its numerical approximation. For instance, if the model under study is Hamiltonian, then the exact flow is symplectic, which motivates the design of symplectic numerical approximation. For practical simulations of Hamiltonian systems, symplectic methods are known to possess very nice properties (see [40]). It is important to combine multiscale techniques to geometric numerical integration. All the problems and equations we intend to deal with will be addressed with a view to preserve intrinsic geometric properties of the exact solutions and/or to approach the asymptotic limit of the system in presence of a small parameter. An example of a numerical method developed by members of the team is the multi-revolution method.
- Quasi-periodic case:

So far, numerical methods have been proposed for the periodic case with single frequency. However, the quasi-periodic case [0] is still misunderstood although many complex problems involve multi-frequencies. Even if the quasi-periodic averaging is doable from a theoretical point of view in the ODE case (see [45]), it is unclear how it can be extended to PDEs. One of the main obstacle being the requirement, usual for ODEs like (4), for $\mathcal{F}$ to be analytic in the periodic variables, an assumption which is clearly impossible to meet in the PDE setting. An even more challenging problem is then the design of numerical methods for this problem.

- extension to stochastic PDEs:
  All these questions will be revisited within the stochastic context. The mathematical study opens the way to the derivation of efficient multiscale numerical schemes for this kind of problems. We believe that the theory is now sufficiently well understood to address the derivation and numerical analysis of multiscale numerical schemes. Multi-revolution composition methods have been recently extended to highly-oscillatory stochastic differential equations The generalization of such multiscale numerical methods to SPDEs is of great interest. The analysis and simulation of numerical schemes for highly-oscillatory nonlinear stochastic Schrödinger equation under diffusion-approximation for instance will be one important objective for us. Finally, an important aspect concerns the quantification of uncertainties in highly-oscillatory kinetic or quantum models (due to an incomplete knowledge of coefficients or imprecise measurement of datas). The construction of efficient multiscale numerical methods which can handle multiple scales as well as random inputs have important engineering applications.

4. Application Domains

4.1. Application domains

The MINGUS project aims at applying the new numerical methods on realistic problems arising for instance in physics of nanotechnology and physics of plasmas. Therefore, in addition to efforts devoted to the design and the analysis of numerical methods, the inherent large size of the problems at hand requires advanced mathematical and computational methods which are hard to implement. Another application is concerned with population dynamics for which the main goal is to understand how the spatial propagation phenomena affect the demography of a population (plankton, parasite fungi, ...). Our activity is mostly at an early stage in the process of transfer to industry. However, all the models we use are physically relevant and all have applications in many areas (ITER, Bose-Einstein condensate, wave turbulence, optical tomography, transport phenomena, population dynamics, $\cdots$). As a consequence, our research aims at reaching theoretical physicists or computational scientists in various fields who have strong links with industrial applications. In order to tackle as realistic physical problems as possible, a fundamental aspect will consist in working on the realization of numerical methods and algorithms which are able to make an efficient use of a large number of processors. Then, it is essential for the numerical methods developed in the MINGuS project to be thought through this prism. We will benefit from the strong expertise of P. Navaro in scientific computing and more precisely on the Selalib software library (see description below). Below, we detail our main applications: first, the modeling and numerical approximation of magnetized plasmas is our major application and will require important efforts in terms of software developments to scale-up our multiscale methods; second, the transport of charged particles in nanostructures has very interesting applications (like graphene material), for which our contributions will mainly focus on dedicated problems; lastly, applications on population dynamics will be dedicated to mathematical modeling and some numerical validations.

[0] replacing t/ε by $t\omega/\varepsilon$ in (4), with $\omega \in \mathbb{R}^d$ a vector of non-resonant frequencies

4.1.1. Plasmas problems

The Selalib (SEmi-LAgrangian LIBrary) software library [0] is a modular library for kinetic and gyrokinetic simulations of plasmas in fusion energy devices. Selalib is a collection of fortran modules aimed at facilitating the development of kinetic simulations, particularly in the study of turbulence in fusion plasmas. Selalib offers basic capabilities and modules to help parallelization (both MPI and OpenMP), as well as pre-packaged simulations. Its main objective is to develop a documented library implementing several numerical methods for the numerical approximation of kinetic models. Another objective of the library is to provide physicists with easy-to-use gyrokinetic solvers. It has been originally developed by E. Sonnendrücker and his collaborators in the past CALVI Inria project, and has played an important role in the activities of the IPL FRATRES. P. Navaro is one of the main software engineer of this library and as such he played an important daily role in its development and its portability on supercomputers. Though Selalib has reached a certain maturity some additional works are needed to make available by the community. There are currently discussions for a possible evolution of Selalib, namely the writing of a new release which will be available for free download. Obviously, the team will be involved in this process. At the scientific level, Selalib is of great interest for us since it provides a powerful tool with which we can test, validate and compare our new methods and algorithms (users level). Besides numerical algorithms the library provides low-level utilities, input-output modules as well as parallelization strategies dedicated to kinetic problems. Moreover, a collection of simulations for typical test cases (of increasing difficulties) with various discretization schemes supplements the library. This library turns out to be the ideal complement of our activities and it will help us to scale-up our numerical methods to high-dimensional kinetic problems. During the last years, several experiments have been successfully performed in this direction (especially with PhD students) and it is important for us that this approach remains throughout. Then, we intend to integrate several of the numerical methods developed by the team within the Selalib library, with the strong help of P. Navaro (contributors level). This work has important advantages: (i) it will improve our research codes (in terms of efficiency but also of software maintenance point of view); (ii) it will help us to promote our research by making our methods available to the research community.

4.1.2. Quantum problems

Nowadays, a great challenge consists in the downscaling at the nanometer scale of electronic components in order to improve speed and efficiency of semiconductor materials. In this task, modeling and numerical simulations play an important role in the determination of the limit size of the nanotransistors. At the nanoscale, quantum effects have to be considered and the Schrödinger equation is prominent equation in this context. In the so-called semiclassical regime or when the transport is strongly confined, the solution endows space-time highly oscillations which are very difficult to capture numerically. An important application is the modeling of charged particles transport in graphene. Graphene is a sheet of carbone made of a single layer of molecule, organized in a bidimensional honeycomb crystal. The transport of charged particles in this structure is usually performed by Dirac equation (which is the relativistic counterpart of the Schrödinger equation). Due to the unusual properties of graphene -at room temperature, electrons moving in graphene behave as massless relativistic particles- physicists and compagnies are nowadays actively studying this material. Here, predicting how the material properties are affected by the uncertainties in the hexagonal lattice structure or in external potentials, is a major issue.

4.1.3. Population dynamics

The main goal is to characterize how spatial propagation phenomena (diffusion, transport, advection, $\cdots$) affect the time evolution of the demography of a population. In collaboration with Y. Lagadeuc (ECOBIO, Rennes), this question has been studied for plankton. In this context, mathematical models have been proposed and it has been shown that the spatial dynamic (in this context, due to the marine current) which is fast compared to demographic scales, can strongly modify the demographic evolution of the plankton. In collaboration with Ecole d'Agronomie de Rennes, a mathematical study on the demography of a parasite fungi of plants has been performed. In this context, the demography is specific: the fungi can proliferate through sexual reproduction or through parthenogenesis. This two ways of reproduction give rise mathematically to quadratic and linear

[0] SELALIB, http://selalib.gforge.inria.fr

growth rates with respect to the population variable. The demography is then coupled with transport (transport of fungi spore by wind). Here, the goal is characterize the propagation of the fungi population by finding travelling waves solutions which are well adapted to describe the evolution of invasive fronts. Moreover, this approach enables to recover with a good agreement realistic examples (infection of ash or banana tree) for which experimental data are available. In these contexts, mathematical models are a powerful tool for biologists since measurements are very complicated to obtain and laboratory experiments hardly reproduce reality. The models derived are multiscale due to the nature of the underlying phenomena and the next step is to provide efficient numerical schemes.

5. New Software and Platforms

5.1. Selalib

SEmi-LAgrangian LIBrary

KEYWORDS: Plasma physics - Semilagrangian method - Parallel computing - Plasma turbulence

SCIENTIFIC DESCRIPTION: The objective of the Selalib project (SEmi-LAgrangian LIBrary) is to develop a well-designed, organized and documented library implementing several numerical methods for kinetic models of plasma physics. Its ultimate goal is to produce gyrokinetic simulations.

Another objective of the library is to provide to physicists easy-to-use gyrokinetic solvers, based on the semi-lagrangian techniques developed by Eric Sonnendrücker and his collaborators in the past CALVI project. The new models and schemes from TONUS are also intended to be incorporated into Selalib.

FUNCTIONAL DESCRIPTION: Selalib is a collection of modules conceived to aid in the development of plasma physics simulations, particularly in the study of turbulence in fusion plasmas. Selalib offers basic capabilities from general and mathematical utilities and modules to aid in parallelization, up to pre-packaged simulations.

- Partners: Max Planck Insitute - Garching - Université de Strasbourg
- Contact: Philippe Helluy
- URL: http://selalib.gforge.inria.fr/

6. New Results

6.1. New Results

Analysis of PDEs and SPDEs

In [17], we prove the nonlinear instability of inhomogeneous steady states solutions to the Hamiltonian Mean Field (HMF) model. We first study the linear instability of this model under a simple criterion by adapting the techniques developed by the authors recently. In a second part, we extend to the inhomogeneous case some techniques developed by the authors recently and prove a nonlinear instability result under the same criterion.

In [24], we consider the non linear wave equation (NLW) on the d-dimensional torus with a smooth nonlinearity of order at least two at the origin. We prove that, for almost any mass, small and smooth solutions of high Sobolev indices are stable up to arbitrary long times with respect to the size of the initial data. To prove this result we use a normal form transformation decomposing the dynamics into low and high frequencies with weak interactions. While the low part of the dynamics can be put under classical Birkhoff normal form, the high modes evolve according to a time dependent linear Hamiltonian system. We then control the global dynamics by using polynomial growth estimates for high modes and the preservation of Sobolev norms for the low modes. Our general strategy applies to any semi-linear Hamiltonian PDEs whose linear frequencies satisfy a very general non resonance condition. The (NLW) equation on a torus is a good example since the standard Birkhoff normal form applies only when $d = 1$ while our strategy applies in any dimension.

In [20], we study semigroups generated by accretive non-selfadjoint quadratic differential operators. We give a description of the polar decomposition of the associated evolution operators as products of a selfadjoint operator and a unitary operator. The selfadjoint parts turn out to be also evolution operators generated by time-dependent real-valued quadratic forms that are studied in details. As a byproduct of this decomposition, we give a geometric description of the regularizing properties of semigroups generated by accretive non-selfadjoint quadratic operators. Finally, by using the interpolation theory, we take advantage of this smoothing effect to establish subelliptic estimates enjoyed by quadratic operators.

In [16], 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.

In [25], we study the Boltzmann equation with external forces, not necessarily deriving from a potential, in the incompressible Navier-Stokes perturbative regime. On the torus, we establish local-in-time, for any time, Cauchy theories that are independent of the Knudsen number in Sobolev spaces. The existence is proved around a time-dependent Maxwellian that behaves like the global equilibrium both as time grows and as the Knudsen number decreases. We combine hypocoercive properties of linearized Boltzmann operators with linearization around a time-dependent Maxwellian that catches the fluctuations of the characteristics trajectories due to the presence of the force. This uniform theory is sufficiently robust to derive the incompressible Navier-Stokes-Fourier system with an external force from the Boltzmann equation. Neither smallness, nor time-decaying assumption is required for the external force, nor a gradient form, and we deal with general hard potential and cutoff Boltzmann kernels. As a by-product the latest general theories for unit Knudsen number when the force is sufficiently small and decays in time are recovered.

In [15], 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.

In [2], we provide in this work a local in time well-posedness result for a quasilinear generalized parabolic Anderson model in dimension two $\partial_t u + \Delta\Pi(u) = g(u)\xi$. The key idea of our approach is a simple transformation of the equation which allows to treat the problem as a semilinear problem. The analysis is done within the setting of paracontrolled calculus.

In [30], we consider the Burgers equation on $H = L^2(0,1)$ perturbed by white noise and the corresponding transition semigroup $P_t D\varphi$. We prove a new formula for $PtD\varphi$ (where $\varphi : H \to \mathbb{R}$ is bounded and Borel) which depends on φ but not on its derivative. Then we deduce some consequences for the invariant measure ν of P_t as its Fomin differentiability and an integration by parts formula which generalises the classical one for gaussian measures.

In [9], we deal with the validity of a large deviation principle for the two-dimensional Navier-Stokes equation, with periodic boundary conditions, 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 ε and $\delta(\varepsilon)$, respectively, with $0 < \varepsilon, \delta(\varepsilon) << 1$. Depending on the relationship between ε and $\delta(\varepsilon)$ we will prove the validity of the large deviation principle in different functional spaces.

In [30], the authors consider the transition semigroup P_t of the Φ^4_2 stochastic quantisation on the torus $\mathbb{T}^2$ and prove the following new estimate

$$|DP_t\varphi(x)\cdot h| \le ct^{-\beta}|h|_{C^{-s}}\|\varphi\|_0(1+|x|_{C^{-s}})^{\gamma},$$

for some α, β, γ, s positive. Thanks to this estimate, we show that cylindrical functions are a core for the corresponding Kolmogorov equation. Some consequences of this fact are discussed in a final remark.

In [32], we consider a particle system with a mean-field-type interaction perturbed by some common and individual noises. When the interacting kernels are sublinear and only locally Lipschitz-continuous, relying on arguments regarding the tightness of random measures in Wasserstein spaces, we are able to construct a weak solution of the corresponding limiting SPDE. In a setup where the diffusion coefficient on the environmental noise is bounded, this weak convergence can be turned into a strong $L^p(\Omega)$ convergence and the propagation of chaos for the particle system can be established. The systems considered include perturbations of the Cucker-Smale model for collective motion.

Numerical schemes

In [7], the asymptotic behavior of the solutions of the second order linearized Vlasov-Poisson system around homogeneous equilibria is derived. It provides a fine description of some nonlinear and multidimensional phenomena such as the existence of Best frequencies. Numerical results for the $1D \times 1D$ and $2D \times 2D$ Vlasov-Poisson system illustrate the effectiveness of this approach.

In [6], we consider the problem of existence and stability of solitary traveling waves for the one dimensional discrete non linear Schrödinger equation (DNLS) with cubic nonlinearity, near the continuous limit. We construct a family of solutions close to the continuous traveling waves and prove their stability over long times. Applying a modulation method, we also show that we can describe the dynamics near these discrete traveling waves over long times.

In [4], we consider the discrete nonlinear Schrödinger equations on a one dimensional lattice of mesh h, with a cubic focusing or defocusing nonlinearity. We prove a polynomial bound on the growth of the discrete Sobolev norms, uniformly with respect to the stepsize of the grid. This bound is based on a construction of higher modified energies.

The efficient numerical solution of many kinetic models in plasma physics is impeded by the stiffness of these systems. Exponential integrators are attractive in this context as they remove the CFL condition induced by the linear part of the system, which in practice is often the most stringent stability constraint. In the literature, these schemes have been found to perform well, e.g., for drift-kinetic problems. Despite their overall efficiency and their many favorable properties, most of the commonly used exponential integrators behave rather erratically in terms of the allowed time step size in some situations. This severely limits their utility and robustness. Our goal in [29] is to explain the observed behavior and suggest exponential methods that do not suffer from the stated deficiencies. To accomplish this we study the stability of exponential integrators for a linearized problem. This analysis shows that classic exponential integrators exhibit severe deficiencies in that regard. Based on the analysis conducted we propose to use Lawson methods, which can be shown not to suffer from the same stability issues. We confirm these results and demonstrate the efficiency of Lawson methods by performing numerical simulations for both the Vlasov-Poisson system and a drift-kinetic model of a ion temperature gradient instability.

In [18], a bracket structure is proposed for the laser-plasma interaction model introduced in the physical literature, and it is proved by direct calculations that the bracket is Poisson which satisfies the Jacobi identity. Then splitting methods in time are proposed based on the Poisson structure. For the quasi- relativistic case, the Hamiltonian splitting leads to three subsystems which can be solved exactly. The conservative splitting is proposed for the fully relativistic case, and three one-dimensional conservative subsystems are obtained. Combined with the splittings in time, in phase space discretization we use the Fourier spectral and finite volume methods. It is proved that the discrete charge and discrete Poisson equation are conserved by our numerical schemes. Numerically, some numerical experiments are conducted to verify good conservations for the charge, energy and Poisson equation.

In [26], the recent advances about the construction of a Trefftz Discontinuous Galerkin (TDG) method to a class of Friedrichs systems coming from linear transport with relaxation are presented in a comprehensive setting. Application to the $2DP_N$ model are discussed, together with the derivation of new high order convergence estimates and new numerical results for the P_1 and P_3 models. More numerical results in dimension 2 illustrate the theoretical properties.

In [8], we are concerned with a formulation of Magnus and Floquet-Magnus expansions for general nonlinear differential equations. To this aim, we introduce suitable continuous variable transformations generated by operators. As an application of the simple formulas so-obtained, we explicitly compute the first terms of the Floquet-Magnus expansion for the Van der Pol oscillator and the nonlinear Schrödinger equation on the torus.

The article [11] is devoted to the construction of numerical methods which remain insensitive to the smallness of the semiclassical parameter for the linear Schrödinger equation in the semiclassical limit. We specifically analyse the convergence behavior of the first-order splitting. Our main result is a proof of uniform accuracy. We illustrate the properties of our methods with simulations.

In [10], we consider the numerical solution of highly-oscillatory Vlasov and Vlasov-Poisson equations with non-homogeneous magnetic field. Designed in the spirit of recent uniformly accurate methods, our schemes remain insensitive to the stiffness of the problem, in terms of both accuracy and computational cost. The specific difficulty (and the resulting novelty of our approach) stems from the presence of a non-periodic oscillation, which necessitates a careful ad-hoc reformulation of the equations. Our results are illustrated numerically on several examples.

In the analysis of highly-oscillatory evolution problems, it is commonly assumed that a single frequency is present and that it is either constant or, at least, bounded from below by a strictly positive constant uniformly in time. Allowing for the possibility that the frequency actually depends on time and vanishes at some instants introduces additional difficulties from both the asymptotic analysis and numerical simulation points of view. This work [13] is a first step towards the resolution of these difficulties. In particular, we show that it is still possible in this situation to infer the asymptotic behaviour of the solution at the price of more intricate computations and we derive a second order uniformly accurate numerical method.

In [12], we introduce a new methodology to design uniformly accurate methods for oscillatory evolution equations. The targeted models are envisaged in a wide spectrum of regimes, from non-stiff to highly-oscillatory. Thanks to an averaging transformation, the stiffness of the problem is softened, allowing for standard schemes to retain their usual orders of convergence. Overall, high-order numerical approximations are obtained with errors and at a cost independent of the regime.

In [1], we present an asymptotic preserving scheme based on a micro-macro decomposition for stochastic linear transport equations in kinetic and diffusive regimes. We perform a mathematical analysis and prove that the scheme is uniformly stable with respect to the mean free path of the particles in the simple telegraph model and in the general case. We present several numerical tests which validate our scheme.

In [22], a splitting strategy is introduced to approximate two-dimensional rotation motions. Unlike standard approaches based on directional splitting which usually lead to a wrong angular velocity and then to large error, the splitting studied here turns out to be exact in time. Combined with spectral methods, the so-obtained numerical method is able to capture the solution to the associated partial differential equation with a very high accuracy. A complete numerical analysis of this method is given in this work. Then, the method is used to design highly accurate time integrators for Vlasov type equations: the Vlasov-Maxwell system and the Vlasov-HMF model. Finally , several numerical illustrations and comparisons with methods from the literature are discussed.

In [23], some exact splittings are proposed for inhomogeneous quadratic differential equations including, for example, transport equations, kinetic equations, and Schrödinger type equations with a rotation term. In this work, these exact splittings are combined with pseudo-spectral methods in space to illustrate their high accuracy and efficiency.

In [14], we develop a new class of numerical schemes for collisional kinetic equations in the diffusive regime. The first step consists in reformulating the problem by decomposing the solution in the time evolution of an equilibrium state plus a perturbation. Then, the scheme combines a Monte Carlo solver for the perturbation with an Eulerian method for the equilibrium part, and is designed in such a way to be uniformly stable with respect to the diffusive scaling and to be consistent with the asymptotic diffusion equation. Moreover, since particles are only used to describe the perturbation part of the solution, the scheme becomes computationally less expensive - and is thus an asymptotically complexity diminishing scheme (ACDS) - as the solution approaches the equilibrium state due to the fact that the number of particles diminishes accordingly. This contrasts with standard methods for kinetic equations where the computational cost increases (or at least does not decrease) with the number of interactions. At the same time, the statistical error due to the Monte Carlo part of the solution decreases as the system approaches the equilibrium state: the method automatically degenerates to a solution of the macroscopic diffusion equation in the limit of infinite number of interactions. After a detailed description of the method, we perform several numerical tests and compare this new approach with classical numerical methods on various problems up to the full three dimensional case.

In [5], 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.

In [28], we introduce a new Monte Carlo method for solving the Boltzmann model of rarefied gas dynamics. The method works by reformulating the original problem through a micro-macro decomposition and successively in solving a suitable equation for the perturbation from the local thermodynamic equilibrium. This equation is then discretized by using unconditionally stable exponential schemes in time which project the solution over the corresponding equilibrium state when the time step is sent to infinity. The Monte Carlo method is designed on this time integration method and it only describes the perturbation from the final state. In this way, the number of samples diminishes during the time evolution of the solution and when the final equilibrium state is reached, the number of statistical samples becomes automatically zero. The resulting method is 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 decreases as the system approaches the equilibrium state. In a last part, we show the behaviors of this new approach in comparison with standard Monte Carlo techniques and in comparison with spectral methods on different prototype problems.

In [27], we consider the three dimensional Vlasov equation with an inhomogeneous, varying direction, strong magnetic field. Whenever the magnetic field has constant intensity, the oscillations generated by the stiff term are periodic. The homogenized model is then derived and several state-of-the-art multiscale methods, in combination with the Particle-In-Cell discretisation, are proposed for solving the Vlasov-Poisson equation. Their accuracy as much as their computational cost remain essentially independent of the strength of the magnetic field. The proposed schemes thus allow large computational steps, while the full gyro-motion can be restored by a linear interpolation in time. In the linear case, extensions are introduced for general magnetic field (varying intensity and direction). Eventually, numerical experiments are exposed to illustrate the efficiency of the methods and some long-term simulations are presented.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

- Contrat with RAVEL (onne year, budget 15000 euros): this is a collaboration with the startup RAVEL on a one-year basis (with possible renewal at the end of the year). The objective is to study the mathematical fondations of artificial intelligence and in particular machine learning algorithms for data anonymized though homomorphic encryption.
  Participants: P. Chartier, M. Lemou and F. Méhats.

- Contract with Cailabs (6 months, budget 3000 euros): This collaboration aims at exploring the possibility of deriving new fiber optics devices based on neural networks architecture.
  Participants: P. Chartier, E. Faou, M. Lemou and F. Méhats.

8. Partnerships and Cooperations

8.1. Regional Initiatives

- M. Lemou and N. Crouseilles are head of the project "MUNIQ" of ENS Rennes. This two-years project (2018-2019) intends to gather multiscale numerical methods and uncertainty quantification techniques. The MINGuS members are P. Chartier, N. Crouseilles, M. Lemou and F. Méhats and colleagues from university of Madison-Wisconsin also belong to this project.

8.2. National Initiatives

8.2.1. ANR

8.2.1.1. MOONRISE: 2015-2019

Participants: François Castella, Philippe Chartier, Nicolas Crouseilles, Mohammed Lemou, Florian Méhats.

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 MINGuS 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.
- Yong Zhang 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. Yong Zhang is now professor at the Tianjin university (China).
- Xiaofei Zhao has been hired as a Postdoc from september 2015 to september 2016 under the supervision of Florian Méhats. Xiaofei Zhao is now postdoc assistant professor in the Wuhan University (China).

8.2.1.2. 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.

8.2.1.3. ADA: 2019-2023

Participant: Arnaud Debussche.

The aim of this project is to treat multiscale models which are both infinite-dimensional and stochastic with a theoretic and computational approach. Multiscale analysis and multiscale numerical approximation for infinite-dimensional problems (partial differential equations) is an extensive part of contemporary mathematics, with such wide topics as hydrodynamic limits, homogenization, design of asymptotic-preserving scheme. Multiscale models in a random or stochastic context have been analysed and computed essentially in finite dimension (ordinary/stochastic differential equations), or in very specific areas, mainly the propagation of waves, of partial differential equations. The technical difficulties of our project are due to the stochastic aspect of the problems (this brings singular terms in the equations, which are difficult to understand with a pure PDE’s analysis approach) and to their infinite-dimensional character, which typically raises compactness and computational issues. Our main fields of investigation are: stochastic hydrodynamic limit (for example for fluids), diffusion-approximation for dispersive equations, numerical approximation of stochastic multiscale equations in infinite dimension. Our aim is to create the new tools - analytical, probabilistic and numerical - which are required to understand a large class of stochastic multiscale partial differential equations. Various modelling issues require this indeed, and are pointing at a new class of mathematical problems that we wish to solve. We also intend to promote the kind of problems we are interested in, particularly among young researchers, but also to recognized experts, via schools, conference, and books.

The partners are ENS Lyon (coordinator J. Vovelle) and ENS Rennes (Coordinator A. Debussche).

8.2.2. Fédération de Recherche : Fusion par Confinement Magnétique

We are involved in the national research multidisciplinary group around magnetic fusion activities. As such, we answer to annual calls.

8.2.3. IPL SURF

A. Debussche and E. Faou are members of the IPL (Inria Project Lab) SURF: Sea Uncertainty Representation and Forecast. Head: Patrick Vidard.

8.2.4. AdT J-Plaff

This AdT started in october 2019 and will be finished in september 2021. An engineer has been hired (Y. Mocquard) to develop several packages in the Julia langage. The J-Plaff is shared with the Fluminance team.

8.3. European Initiatives

8.3.1. Collaborations in European Programs, Except FP7 & H2020

Program: Eurofusion

Project acronym: MAGYK

Project title:

Duration: january 2019-december 2020

Coordinator: E. Sonnendrücker

Other partners: Switzerland, Germany, France, Austria, Finland.

Abstract: This proposal is aimed at developing new models and algorithms that will be instrumental in enabling the efficient and reliable simulation of the full tokamak including the edge and scrape-off layer up to the wall with gyrokinetic or full kinetic models. It is based on a collaboration between applied mathematicians and fusion physicists that has already been very successful in a previous enabling research project and brings new ideas and techniques into the magnetic fusion community. New modelling and theoretical studies to extend the modern gyrokinetic theory up to the wall including boundary conditions will be addressed, and the limits of gyrokinetics will be assessed. New multiscale methods will enable to efficiently and robustly separate time scales, which will on the one hand make gyrokinetic codes more efficient and on the other hand enable full implicit kinetic simulations. Difficult algorithmic issues for handling the core to edge transition, the singularities at the O- and X-points will be addressed. And finally, pioneering work based on recent (deep) machine learning techniques will be performed, on the one hand to automatically identify a Partial Differential Equation (PDE) from the data, which can be used for verification and sensitivity analysis purposes, and on the other hand to develop reduced order models that will define a low- cost low-fidelity model based on the original high-fidelity gyrokinetic or kinetic model that can be used for parameter scans and uncertainty quantification.

8.4. International Initiatives

8.4.1. Inria Associate Teams Not Involved in an Inria International Labs

8.4.1.1. ANTIpODE

Title: Asymptotic Numerical meThods for Oscillatory partial Differential Equations with uncertainties

International Partner (Institution - Laboratory - Researcher):

University of Wisconsin-Madison, USA (United States)

Start year: 2018

See also: https://team.inria.fr/antipode/

The proposed associate team assembles the Inria team IPSO and the research group led by Prof. Shi Jin from the Department of Mathematics at the University of Wisconsin, Madison. The main scientific objective of ANTIpODE consists in marrying uniformly accurate and uncertainty quantification techniques for multi-scale PDEs with uncertain data. Multi-scale models, as those originating e.g. from the simulation of plasma fusion or from quantum models, indeed often come with uncertainties. The main scope of this proposal is thus (i) the development of uniformly accurate schemes for PDEs where space and time high oscillations co-exist and (ii) their extension to models with uncertainties. Applications to plasmas (Vlasov equations) and graphene (quantum models) are of paramount importance to the project.

8.4.2. Inria International Partners

8.4.2.1. Informal International Partners

The members of MINGuS have several interactions with the following partners

- Europe: University of Geneva (Switzerland), University of Jaume I (Spain), University of Basque Country (Spain), University of Innsbruck (Austria), University of Ferrare (Italy), Max Planck Institute (Germany), SNS Pisa (Italy)
- USA: Georgia Tech, University of Maryland, University of Wisconsin, NYU
- Asia: Chinese Academy of Science (China), University of Wuhan (China), shanghai jiao tong university (China), National University of Singapore (Singapore)

8.4.3. Participation in Other International Programs

- SIMONS project. Erwan Faou is one of the Principal investigators of the Simons Collaboration program *Wave Turbulence*. Head: Jalal Shatah (NYU).

8.5. International Research Visitors

8.5.1. Visits of International Scientists

- Fernando Casas (University of Jaume I, Spain) was invited in the MINGuS team during 6 months (september 2018 to february 2019), funded by the Labex (CHL) Center Henri Lebesgue.
- Yingzhe Li (University of Chinese Academy of Sciences, China) is visiting the IRMAR laboratory during one year (March 2019-February 2020) thanks to a chinese grant. He is currently a PhD student advised by Yajuan Sun, professor at CAS.
- Xiaofei Zhao (University of Wuhan, China) was invited in the MINGuS team during 2 weeks (july 2019).
- Yoshio Tsutsumi (Kyoto University, Japan) was invited in the IRMAR laboratory during 2 months (october-november 2019).

8.5.1.1. Internships

- G. Barrué: Master 2 internship, A. Debussche.
- Q. Chauleur: Master 2 internship, R. Carles (CNRS, Rennes) and E. Faou.
- U. Léauté: Master 1 internship, B. Boutin (University Rennes I and N. Crouseilles).
- A. V. Tuan: Master 2 internship, M. Lemou and F. Méhats.

8.5.2. Visits to International Teams

8.5.2.1. Sabbatical programme

P. Chartier was on a sabbatical visit from the 1st of February to the 30th of September 2019 at the University of the Basque Country, Spain.

8.5.2.2. Research Stays Abroad

- P. Chartier was invited by G. Vilmart, University of Geneva, Geneva, Switzerland, January 2019.
- P. Chartier was invited by F. Casas at the university of Jaume I, Castellon, Spain, July 2019.
- P. Chartier was Invited by Q. Li at the university of Wisconsin, Madison, USA, September 2019.
- P. Chartier was Invited by M. Tao at Georgia Tech, Atlanta, USA, August 2019.
- A. Debussche was invited by G. Da Prato at Scuola Normale Superiore, Pise, Italy, April 2019.
- E. Faou was a participant of the Semester *Geometry, compatibility and structure preservation in computational differential equations*, Isaac Newton Institute, Cambridge, UK (3 months stay, September-December 2019).
- M. Lemou was invited by J. Joudioux and L. Anderson, at the Albert Einstein Institute, Golm, Germany, February 2019.
- M. Lemou was invited by A. M. M. Luz at the Universidade Federal Fluminense, Rio de Janeiro, Brazil, April 2019.
- M. Lemou was invited by S. Jin at Shanghai Jiao Tong University, Shanghai, China, April 2019.
- M. Lemou was invited by J. Ben-Artzi at the university of Cardiff, Cardiff, UK, May 2019.
- M. Lemou was invited by G. Vilmart, University of Geneva, Geneva, Switzerland, January 2019.
- M. Lemou was Invited by Q. Li at the university of Wisconsin, Madison, USA, September 2019.
- M. Lemou was Invited by M. Tao at Georgia Tech, Atlanta, USA, August 2019.

- F. Méhats was invited by A. de la Luz at the Universidade Federal Fluminense, Rio de Janeiro, Brazil, April 2019.
- F. Méhats was invited by G. Vilmart, University of Geneva, Geneva, Switzerland, January 2019.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Organisation

9.1.1.1. Member of the Organizing Committees

- F. Castella organized the MINGuS team meeting, Dinard, december 2019. [15 participants]
- P. Chartier, N. Crouseilles, M. Lemou and F. Méhats organized a workshop on "Asymptotic methods and numerical approximations of multi-scale evolution problems, and uncertainty quantification", ENS Rennes, May 2019. [30 participants]
- N. Crouseilles is co-organizer of the weekly seminar "Mathematic and applications" at ENS Rennes.
- P. Chartier and M. Lemou co-organized (with M. Thalhammer, university of Innsbruck) the mini-symposium "Advanced numerical methods for differential equations" in the ICIAM 2019 conference, Valencia, Spain, July 2019.
- F. Méhats co-organized (with W. Bao, National University of Singapore) the mini-symposium "Multiscale methods and analysis for oscillatory PDEs" in the Scicade conference, Innsbruck, Austria, July 2019.

9.1.1.2. Member of the Conference Program Committees

- N. Crouseilles was member of the scientific committee of the SMAI-19 Conference, June 2019.
- E. Faou was member of the scientific committee of the Scicade Conference, July 2019.

9.1.2. Journal

9.1.2.1. Member of the Editorial Boards

- P. Chartier is member of the editorial board of "Mathematical Modelling and Numerical Analysis" (M2AN).
- A. Debussche is editor in chief of Stochastic Partial Differential Equations: analysis and computations (2013-).
- A. Debussche is member of the editorial board of the collection "Mathématiques & Applications".
- A. Debussche is associate editor of Differential and Integral Equations (2002-19).
- A. Debussche is associate editor of Potential Analysis (2011-2019).
- A. Debussche is associate editor of Journal of Evolution Equation (2014-).
- A. Debussche is associate editor of Applied Mathematics & Optimization, SIAM JUQ (2016-2019).
- M. Lemou is member of the editorial committee of "Communications in Mathematical Sciences" (CMS).

9.1.2.2. Reviewer - Reviewing Activities

The members of the MINGuS team are reviewers for almost all the journals in which they publish (SIAM, JCP, CPDE, CMP, ARMA, JSP, JSC, JMAA, ANM, JCAM, NMPDE, Numer. Math., $\cdots$).

9.1.3. Invited Talks

- J. Bernier gave a talk in the seminar of Columbia university, New York, USA, June 2019.
- P. Chartier gave a talk in the workshop *Nonlinear Evolution Equations: Analysis and Numerics*, organized by M. Hochbruck, H. Koch, S.-J. Oh, and A. Ostermann, Oberwolfach, Germany, February 2019.
- P. Chartier gave a talk in the workshop HaLu, Gran Sasso Science Institute (GSSI) School of Advanced Studies, L'Aquila, Italy, June 2019.
- P. Chartier gave a talk in the ICIAM conference, Valencia, Spain, July 2019.
- P. Chartier gave a talk in the Scicade conference, University of Innsbruck, Austria, July 2019.
- N. Crouseilles gave a talk at the workshop *Quantum and Kinetic Transport*, Shanghai, China, April 2019.
- N. Crouseilles gave a talk in the workshop *Numerical Methods for Multiscale Models arising in Physics and Biology*, University of Nantes, France, June 2019.
- N. Crouseilles gave a talk in the Scicade conference, University of Innsbruck, Austria, July 2019.
- A. Debussche gave a talk in the conference *Partial Differential Equations: from theory to applications*, Nancy, France, March 2019.
- A. Debussche gave a talk in the workshop *Numerical Methods for SPDE: 20 Successful Years and Future Challenges*, Reims, France, June 2019.
- A. Debussche gave a mini-course in the workshop *PROPAL : propagation d'ondes en milieux aléatoires*, Mittag-Leffler Institute, Stockholm, Sweden, May 2019
- A. Debussche gave a talk in the workshop *Recents Trends in Stochastic Analysis and SPDEs*, University of Pisa, Italy, July 2019.
- A. Debussche gave a talk in the workshop *Touch down of Stochastic Analysis*, University of Bielefeld, Germany, September 2019.
- A. Debussche gave a talk in the conference *Challenges and New Perspectives in Mathematics*, Hassan II Academy of Sciences and Technology, Morocco, November 2019.
- A. Debussche gave a talk in the conference *Paths between probability, PDEs and physics*, Imperial College, July 2019.
- E. Faou gave a talk in the workshop *The future of structure-preserving algorithms*, ICMS, Edinburgh, UK, October 2019.
- E. Faou gave a talk in the Analysis seminar, CMS University of Cambridge, UK, October 2019.
- E. Faou gave a talk in the *Plasma day* session, at the Isaac Newton Institute, Cambridge, UK, October 2019.
- E. Faou gave a Colloquium talk at the University of Bielefeld, Germany, April 2019.
- E. Faou gave a talk in the workshop *Dynamics of nonlinear dispersive PDEs*, La Thuile, Italy, February 2019.
- E. Faou gave a talk in the workshop *Nonlinear Evolution Equations: Analysis and Numerics*, organized by M. Hochbruck, H. Koch, S.-J. Oh, and A. Ostermann, Oberwolfach, Germany, February 2019.
- Y. Li gave a talk in the Scicade conference, University of Innsbruck, July 2019.
- Y. Li gave a talk in the NumKin conference, Max Planck Institute, Garching, October 2019.
- M. Lemou gave a talk at the Albert Einstein Institute, Golm, Germany, February 2019.
- M. Lemou gave a talk at the workshop *Quantum and Kinetic Transport*, Shanghai, China, April 2019.
- M. Lemou gave a talk at the university of Cardiff seminar, Cardiff, UK, May 2019.

- M. Lemou gave a talk in the ICIAM conference, Valencia, Spain, July 2019.
- M. Lemou gave a talk in the Scicade conference, Innsbruck, Austria, July 2019.
- M. Lemou gave a talk at the university of Wisconsin seminar, Madison, USA, September 2019.
- M. Lemou gave a talk at the Georgia Tech seminar, Atlanta, USA, September 2019.
- M. Lemou gave a talk in the workshop *Recent Progress and Challenge in Quantum and Kinetic Problems*, Singapore, October 2019.
- J. Massot gave a talk in the NumKin conference, Max Planck Institute, Garching, October 2019.
- F. Méhats gave a talk in the workshop *Recent Progress and Challenge in Quantum and Kinetic Problems*, Singapore, October 2019.
- F. Méhats gave a talk in the NumKin conference, Max Planck Institute, Garching, October 2019.
- P. Navaro participated at the conference JuliaCon 2019, Baltimore, USA, August 2019.
- A. Rosello gave a talk in the conference *Paths between probability, PDEs and physics*, Imperial College, July 2019.

9.1.4. Scientific Expertise

- N. Crouseilles was member of the committee of the Blaise Pascal prize (GAMNI-SMAI), 2019.
- N. Crouseilles was member of the committee of the best PhD talk in the Scicade conference, 2019.
- N. Crouseilles was member of the committee of the expert reviewers for the European Doctoral programme of the University of Innsbruck.
- A. Debussche was reviewer for ERC projects.
- E. Faou was member of the committee of the PhD prize SMAI-GAMNI, 2019. 2019).

9.1.5. Research Administration

- F. Castella is member of the UFR mathématiques council, University Rennes 1.
- N. Crouseilles is responsible of Fédération Recherche Fusion for the University of Rennes I.
- N. Crouseilles is member of the IRMAR laboratory council, University Rennes 1.
- N. Crouseilles is member of the scientific council of ENS Rennes (until september 2019).
- 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 is scientific vice-deputy and international relations of ENS Rennes.
- A. Debussche is co-director of the Henri Lebesgue Center (Excellence laboratory of the program investissement d'avenir).
- A. Debussche is vice-head of the Lebesgue agency for Mathematic and Innovation.
- E. Faou is co-director of the Henri Lebesgue Center (Excellence laboratory of the program investissement d'avenir).
- E. Faou is member of the Scientific Council of the Pôle Universitaire Léonard de Vinci.
- M. Lemou is the head of the IRMAR team "Analyse numérique" composed of 48 members.
- M. Lemou is member of the scientific council of ENS Rennes.
- M. Lemou is member of the scientific council of the Henri Lebesgue Center.
- P. Navaro is member of the national network "calcul" http://calcul.math.cnrs.fr. This network is well known for interdisciplinarity of CNRS dedicated to technological aspects of scientific computing (programming, optimization, architectures, $\cdots$).

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Master :

- F. Castella, Numerical methods for ODEs and PDEs, 60 hours, Master 1, University of Rennes.
- N. Crouseilles, Numerical methods for PDEs, 24 hours, Master 1, ENS Rennes.
- E. Faou, Normal forms, 24 hours, Master 2, University of Rennes.
- M. Lemou, Numerical methods for kinetic equations, 18 hours, Master 2, University of Rennes.
- M. Lemou, elliptic PDEs, 36 hours, Master 1, University of Rennes.
- P. Navaro, Python courses, Master 2 Smart Data, ENSAI.
- P. Navaro, Scientific computing tools for big data, Master 2, University of Rennes.

9.2.2. Supervision

PhD : G. Barrué, Approximation diffusion pour des équations dispersives, University of Rennes I, started in september 2019, A. Debussche.

PhD : J. Bernier, Study of some perturbation of equations which involve symmetries: resonancy and stability, University of Rennes I, defended in july 2019, E. Faou and N. Crouseilles.

PhD : Q. Chauleur, Equation de Vlasov singulière et équations reliées, University of Rennes I, started in september 2019, R. Carles (CNRS, Rennes) and E. Faou.

PhD: Y. Li (Chinese Academy of Sciences), Structure preserving methods for Vlasov equations, march 2019-february 2020, Y. Sun (Chinese Academy of Sciences) and N. Crouseilles.

PhD in progress : J. Massot, Exponential methods for hybrid kinetic models, started in october 2018, N. Crouseilles.

PhD in progress : A. Rosello, Approximation-diffusion pour des équations cinétiques pour les modèles de type spray, started in september 2016, A. Debussche and J. Vovelle (CNRS, Lyon).

PhD in progress : L. Trémant, Asymptotic analysis methods and numerical of dissipative multi-scale models: ODE with central manifold and kinetic models, started in october 2018, P. Chartier and M. Lemou.

9.2.3. Juries

- F. Castella was referee for the PhD thesis of H. Moundoyi (Laboratoire de biologie marine de Roscoff, France), supervised by P. Cormier and B. Sarels.
- F. Castella was referee for the PhD thesis of F. Patout (ENS Lyon, France), supervised by V. Calvez and J. Garnier.
- N. Crouseilles was referee for the PhD thesis of B. Fedele (University Toulouse 3, France), supervised by C. Negulescu and M. Ottaviani (CEA).
- A. Debussche was referee for the PhD of E. Altmann (Sorbonne Université, France), supervised by L. Zambotti.
- A. Debussche was member of the defense committee of the PhD of B. Kouegou Kamen (Aix-Marseille Université, France), supervised by E. Pardoux.
- A. Debussche was referee for the PhD of T. Yeo (Aix-Marseille université, France) supervised by E. Pardoux.
- E. Faou was referee for the Habilitation degree of Karolina Kropielnicka (Univ. Gdansk, Poland).
- M. Lemou was referee of the PhD thesis of X. Li (University Paris-Dauphine, France), supervised by J. Dolbeault.
- F. Méhats was referee of the PhD thesis of T. Dolmaire (university Paris Diderot, France), supervised by L. Desvillettes and I. Gallagher.

9.3. Popularization

9.3.1. Internal or external Inria responsibilities

- F. Castella was member of the HCERES committee for the evaluation of the laboratory "Mathématiques et Informatique pour la Complexité et les Systèmes (MICS)", CentraleSupelec, Gif-sur-Yvette.
- P. Chartier was member of the hiring committee CR2-Inria (Bordeaux)
- P. Chartier was member of the hiring committee for the Inria promotion DR1-DR0.
- N. Crouseilles is member of the Inria Evaluation Committe (2019-2023).
- N. Crouseilles was member of the Inria hiring committee for the following Inria promotions: CRHC, DR2-DR1, DR1-DR0, DR0-DR02.
- N. Crouseilles is member of the hiring committee of the professor position, ENS Rennes.
- E. Faou was member of the CNU 26 until the summer 2019.
- E. Faou was member of the HCERES committee for the evaluation of the Mathematics Institute of Toulouse, University Paul Sabatier.
- E. Faou is AMIES correspondent (Agency for Interaction in Mathematics with Business and Society) for Inria Rennes Bretagne atlantique and IRMAR.

9.3.2. Interventions

- N. Crouseilles: participation to high school students internship at IRMAR laboratory (one week), June 2019.
- N. Crouseilles: interview by a first year student of University Rennes I (in order to inform the different ways to become Inria researcher).
- J. Massot: talk at the N. Mandela high school (terminal S classes) about the links between astronomy and mathematics, April 2019.
- J. Massot: participation to high school students internship at IRMAR laboratory (one week), June 2019.
- P. Navaro: participation to the Julia day, Lyon, France, January 2019.
- P. Navaro: participation to the Julia day, Nantes, France, June 2019.
- A. Rosello: participation to "MATHC2+", June 2019, ENS Rennes.
- L. Trémant: participation to "Maths en Jean", June 2019, University of Rennes.

9.3.3. Internal action

- J. Massot: writing of a Python library *ponio* (Python Objects for Numerical IntegratOr) https://pypi.org/project/ponio/0.1/
- P. Navaro: Python training at the IRMAR laboratory.
- P. Navaro: Julia training at the IRMAR laboratory.
- P. Navaro: R training at the Finist'R internal workshop, Roscoff, France, August 2019.
- P. Navaro, Python courses, within the exchange program between ENSAI and Hong Kong university.

10. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] N. AYI, E. FAOU.*Analysis of an asymptotic preserving scheme for stochastic linear kinetic equations in the diffusion limit*, in "SIAM/ASA Journal on Uncertainty Quantification", 2019, vol. 7, n^o 2, p. 760-785, https://arxiv.org/abs/1803.06130 [*DOI :* 10.1137/18M1175641], https://hal.archives-ouvertes.fr/hal-01734515

[2] I. BAILLEUL, A. DEBUSSCHE, M. HOFMANOVA.*Quasilinear generalized parabolic Anderson model*, in "Stochastics and Partial Differential Equations: Analysis and Computations", 2019, vol. 7, n^o 1, p. 40-63, https://arxiv.org/abs/1610.06726 [*DOI :* 10.1007/S40072-018-0121-1], https://hal.archives-ouvertes.fr/hal-01389489

[3] K. BELABAS, D. BERNARDI, B. PERRIN-RIOU.*Polygones fondamentaux d'une courbe modulaire*, in "Publications Mathématiques de Besançon : Algèbre et Théorie des Nombres", 2019, p. 1-37, forthcoming, https://hal.archives-ouvertes.fr/hal-01828430

[4] J. BERNIER.*Bounds on the growth of high discrete Sobolev norms for the cubic discrete nonlinear Schrödinger equations on $h\mathbb{Z}$*, in "Discrete and Continuous Dynamical Systems - Series A", 2019, vol. 39, n^o 6, p. 3179-3195, https://arxiv.org/abs/1805.02468 [*DOI :* 10.3934/DCDS.2019131], https://hal.archives-ouvertes.fr/hal-01785953

[5] J. BERNIER.*Optimality and resonances in a class of compact finite difference schemes of high order*, in "Calcolo", 2019, vol. 56, n^o 2, article 12 https://arxiv.org/abs/1710.02953, forthcoming [*DOI :* 10.1007/S10092-019-0309-4], https://hal.archives-ouvertes.fr/hal-01612326

[6] J. BERNIER, E. FAOU.*Existence and stability of traveling waves for discrete nonlinear Schroedinger equations over long times*, in "SIAM Journal on Mathematical Analysis", 2019, vol. 51, n^o 3, p. 1607–1656, https://arxiv.org/abs/1805.03578 [*DOI :* 10.1137/18M1186484], https://hal.archives-ouvertes.fr/hal-01788398

[7] J. BERNIER, M. MEHRENBERGER.*Long-time behavior of second order linearized Vlasov-Poisson equations near a homogeneous equilibrium*, in "Kinetic and Related Models ", 2020, vol. 13, n^o 1, p. 129-168, https://arxiv.org/abs/1903.08374 [*DOI :* 10.3934/KRM.2020005], https://hal.archives-ouvertes.fr/hal-02070138

[8] F. CASAS, P. CHARTIER, A. MURUA.*Continuous changes of variables and the Magnus expansion*, in "Journal of Physics Communications", 2019, p. 1-14, forthcoming [*DOI :* 10.1088/2399-6528/AB42C1], https://hal.inria.fr/hal-02393566

[9] S. CERRAI, A. DEBUSSCHE.*Large deviations for the two-dimensional stochastic Navier-Stokes equation with vanishing noise correlation*, in "Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques", 2019, vol. 55, n^o 1, p. 211-236, https://arxiv.org/abs/1603.02527 [*DOI :* 10.1214/17-AIHP881], https://hal.archives-ouvertes.fr/hal-01942681

[10] P. CHARTIER, N. CROUSEILLES, M. LEMOU, F. MÉHATS, X. ZHAO.*Uniformly accurate methods for Vlasov equations with non-homogeneous strong magnetic field*, in "Mathematics of Computation", 2019, vol. 88, n^o 320, p. 2697-2736 [*DOI :* 10.1090/MCOM/3436], https://hal.inria.fr/hal-01703477

[11] P. CHARTIER, L. LE TREUST, F. MÉHATS.*Uniformly accurate time-splitting methods for the semiclassical linear Schrödinger equation*, in "ESAIM: Mathematical Modelling and Numerical Analysis", 2019, vol. 53, n^o 2, p. 443-473, https://arxiv.org/abs/1601.04825 [*DOI :* 10.1051/M2AN/2018060], https://hal.archives-ouvertes.fr/hal-01257753

[12] P. CHARTIER, M. LEMOU, F. MÉHATS, G. VILMART.*A new class of uniformly accurate numerical schemes for highly oscillatory evolution equations*, in "Foundations of Computational Mathematics", 2019, p. 1-29, forthcoming [*DOI :* 10.1007/S10208-019-09413-3], https://hal.archives-ouvertes.fr/hal-01666472

[13] P. CHARTIER, M. LEMOU, F. MÉHATS, G. VILMART.*Highly-oscillatory problems with time-dependent vanishing frequency*, in "SIAM Journal on Numerical Analysis", 2019, vol. 57, n^{o} 2, p. 925–944, https://arxiv.org/abs/1807.07835 [*DOI :* 10.1137/18M1203456], https://hal.inria.fr/hal-01845614

[14] A. CRESTETTO, N. CROUSEILLES, G. DIMARCO, M. LEMOU.*Asymptotically complexity diminishing schemes (ACDS) for kinetic equations in the diffusive scaling*, in "Journal of Computational Physics", 2019, vol. 394, p. 243-262 [*DOI :* 10.1016/J.JCP.2019.05.032], https://hal.archives-ouvertes.fr/hal-01849671

[15] A. DEBUSSCHE, J. MARTIN.*Solution to the stochastic Schrödinger equation on the full space*, in "Nonlinearity", 2019, vol. 32, n^{o} 4, p. 1147-1174, https://arxiv.org/abs/1707.06431 [*DOI :* 10.1088/1361-6544/AAF50E], https://hal.archives-ouvertes.fr/hal-01579115

[16] M. FONTAINE, M. LEMOU, F. MÉHATS.*Stable Ground States for the HMF Poisson Model*, in "Annales de l'Institut Henri Poincaré (C) Non Linear Analysis", 2019, vol. 36, n^{o} 1, p. 217-255, https://arxiv.org/abs/1709.02234 [*DOI :* 10.1016/J.ANIHPC.2018.05.002], https://hal.archives-ouvertes.fr/hal-01582008

[17] M. LEMOU, A. M. M. LUZ, F. MÉHATS.*Nonlinear instability of inhomogeneous steady states solutions to the HMF Model*, in "Journal of Statistical Physics", 2019, https://arxiv.org/abs/1902.09785 , forthcoming [*DOI :* 10.1007/S10955-019-02448-4], https://hal.archives-ouvertes.fr/hal-02048776

[18] Y. LI, N. CROUSEILLES, Y. SUN.*Numerical simulations of Vlasov-Maxwell equations for laser plasmas based on Poisson structure*, in "Journal of Computational Physics", 2020, p. 1-34, https://hal.inria.fr/hal-02391668

Conferences without Proceedings

[19] A. MULLER-GUEUDIN, A. DEBUSSCHE, A. CRUDU.*Modeling of gene regulation networks by deterministic processes by pieces*, in "Journée de la Fédération Charles Hermite", Vandoeuvre-les-Nancy, France, June 2019, p. 1-37, https://hal.archives-ouvertes.fr/hal-02360992

Other Publications

[20] P. ALPHONSE, J. BERNIER.*Polar decomposition of semigroups generated by non-selfadjoint quadratic differential operators and regularizing effects*, January 2020, https://arxiv.org/abs/1909.03662 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02280971

[21] M. ANTOÑANA, P. CHARTIER, J. MAKAZAGA, A. MURUA.*Global time-regularization of the gravitational N -body problem*, January 2020, working paper or preprint, https://hal.inria.fr/hal-02431607

[22] J. BERNIER, F. CASAS, N. CROUSEILLES.*Splitting methods for rotations: application to vlasov equations*, July 2019, working paper or preprint, https://hal.inria.fr/hal-02178952

[23] J. BERNIER, N. CROUSEILLES, Y. LI.*Exact splitting methods for kinetic and Schrödinger equations*, December 2019, https://arxiv.org/abs/1912.13221 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02425605

[24] J. BERNIER, E. FAOU, B. GREBERT.*Long time behavior of the solutions of NLW on the d-dimensional torus*, September 2019, https://arxiv.org/abs/1906.05107 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02151338

[25] M. BRIANT, A. DEBUSSCHE, J. VOVELLE.*The Boltzmann equation with an external force on the torus: Incompressible Navier-Stokes-Fourier hydrodynamical limit*, June 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02150286

[26] C. BUET, B. DESPRÉS, G. MOREL.*TDG method for Friedrichs systems and the P N model in 2D*, November 2019, working paper or preprint, https://hal.sorbonne-universite.fr/hal-02372279

[27] P. CHARTIER, N. CROUSEILLES, M. LEMOU, F. MÉHATS, X. ZHAO.*Uniformly accurate methods for three dimensional Vlasov equations under strong magnetic field with varying direction*, July 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02179534

[28] A. CRESTETTO, N. CROUSEILLES, G. DIMARCO, M. LEMOU.*A new deviational Asymptotic Preserving Monte Carlo method for the homogeneous Boltzmann equation*, December 2019, working paper or preprint, https://hal.inria.fr/hal-02413232

[29] N. CROUSEILLES, L. EINKEMMER, J. MASSOT.*Exponential methods for solving hyperbolic problems with application to kinetic equations*, October 2019, https://arxiv.org/abs/1910.12720 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02321916

[30] G. DA PRATO, A. DEBUSSCHE.*Estimate for P t D for the stochastic Burgers equation*, November 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02347398

[31] G. DA PRATO, A. DEBUSSCHE.*Gradient Estimates and Maximal Dissipativity for the Kolmogorov Operator in* Φ_2^4, December 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02390677

[32] A. ROSELLO.*Weak and strong mean-field limits for stochastic Cucker-Smale particle systems*, June 2019, https://arxiv.org/abs/1905.02499 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02120106

References in notes

[33] C. BIRDSALL, A. LANGDON.*Plasmas physics via computer simulations*, Taylor and Francis, New York, 2005

[34] A. BRIZARD, T. HAHM.*Foundations of nonlinear gyrokinetic theory*, in "Reviews of Modern Physics", 2007, vol. 79

[35] J. CARR.*Applications of Centre Manifold Theory*, in "Applied Mathematical Sciences Series", 1981, vol. 35

[36] P. CHARTIER, N. CROUSEILLES, M. LEMOU, F. MÉHATS.*Uniformly accurate numerical schemes for highly-oscillatory Klein-Gordon and nonlinear Schrödinger equations*, in "Numer. Math.", 2015, vol. 129, p. 513–536

[37] P. CHARTIER, A. MURUA, J. SANZ-SERNA.*Higher-order averaging, formal series and numerical integration III: error bounds*, in "Foundation of Comput. Math.", 2015, vol. 15, p. 591–612

[38] A. DEBUSSCHE, J. VOVELLE.*Diffusion limit for a stochastic kinetic problem*, in "Commun. Pure Appl. Anal.", 2012, vol. 11, p. 2305–2326

[39] E. FAOU, F. ROUSSET.*Landau damping in Sobolev spaces for the Vlasov-HMF model*, in "Arch. Ration. Mech. Anal.", 2016, vol. 219, p. 887–902

[40] 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

[41] S. JIN, H. LU.*An Asymptotic-Preserving stochastic Galerkin method for the radiative heat transfer equations with random inputs and diffusive scalings*, in "J. Comp. Phys.", 2017, vol. 334, p. 182–206

[42] M. LEMOU, F. MÉHATS, P. RAPHAËL.*Orbital stability of spherical galactic models*, in "Invent. Math.", 2012, vol. 187, p. 145–194

[43] C. MOUHOT, C. VILLANI.*On Landau damping*, in "Acta Math.", 2011, vol. 207, p. 29–201

[44] S. NAZARENKO.*Wave turbulence*, Springer-Verlag, 2011

[45] L. PERKO.*Higher order averaging and related methods for perturbed periodic and quasi-periodic systems*, in "SIAM J. Appl. Math.", 1969, vol. 17, p. 698–724

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

École normale supérieure de Rennes

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Distributed Systems and middleware

Table of contents

Project-Team Myriads

Creation of the Team: 2010 January 01, updated into Project-Team: 2012 January 01

Keywords:

Computer Science and Digital Science:

A1.1.9. - 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
A1.3.2. - Mobile distributed systems
A1.3.4. - Peer to peer
A1.3.5. - Cloud
A1.3.6. - Fog, Edge
A1.6. - Green Computing
A2.1.7. - Distributed programming
A2.2.5. - 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
A2.6.4. - Ressource management
A3.1.2. - Data management, quering and storage
A3.1.3. - Distributed data
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)
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
B8.3. - Urbanism and urban planning
B8.5. - Smart society
B9.1. - Education
B9.1.1. - E-learning, MOOC
B9.1.2. - Serious games
B9.5.1. - Computer science
B9.7. - Knowledge dissemination
B9.7.1. - Open access
B9.7.2. - Open data
B9.8. - Reproducibility
B9.9. - Ethics
B9.10. - Privacy

1. Team, Visitors, External Collaborators

Research Scientist
Anne-Cécile Orgerie [CNRS, Researcher]

Faculty Members
Guillaume Pierre [Team leader, Univ de Rennes I, Professor]
Martin Quinson [Ecole normale supérieure de Rennes, Professor, HDR]
Cédric Tedeschi [Univ de Rennes I, Associate Professor, HDR]
Nikos Parlavantzas [INSA Rennes, Associate Professor]
Jean-Louis Pazat [INSA Rennes, Professor, HDR]
Marin Bertier [INSA Rennes, Associate Professor]

External Collaborators
Anne Blavette [CNRS]
Julie Montégu [Univ de Rennes I]
Louis Rilling [DGA]

Technical Staff
Ehsan Azimi [Inria, since Dec 2019]
Sébastien Philippot [Inria]

PhD Students
Adrien Gougeon [Ecole Normale Supérieure Rennes, since Sep 2019]

Ali Fahs [Univ de Rennes I]
Amir Teshome Wonjiga [Inria, until Feb 2019]
Arif Ahmed [Univ de Rennes I]
Clement El Baz [Inria]
Davaadorj Battulga [U Hopper]
Dorra Boughzala [Inria, since Jul 2019]
Genc Tato [Inria, until Feb 2019]
Hamidreza Arkian [Univ de Rennes I]
Loic Guegan [Ecole normale supérieure de Rennes]
Mehdi Belkhiria [Univ de Rennes I]
Mozhdeh Farhadi [U Hopper]
Mulugeta Tamiru [Elastisys]
Paulo Rodrigues de Souza Junior [Univ de Rennes I]
The Anh Pham [Inria]
Yasmina Bouizem [Rennes Métropole]

Post-Doctoral Fellows
Kameswar Rao Vaddina [Inria, since Nov 2019]
Benjamin Camus [CNRS, until Sep 2019]
Aymen Jlassi [Univ de Rennes I]
Millian Poquet [Inria, until Jun 2019]

Visiting Scientists
Gene Cooperman [Khoury College of Computer, May 2019]
Ayan Mondal [Indian Institute of Technology Kharagpur, since Jun 2019 until Sep 2019]
Deborah Agarwal [Berkeley Lab]

Administrative Assistant
Caroline Tanguy [Univ de Rennes I, since Dec 2019]

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 [0] 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.

[0] 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 resources are offered by companies on a "pay-as-you-go" basis. There is no more need for large 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 *Linkedln* 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 over 12 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 and fog computing 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 fogs and 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.

We expect to integrate the results of these works in the Discovery initiative [33] 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 fog computing architectures

Fog computing 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, fog 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.

Fog computing platforms 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 multi-cloud 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

The ICT (Information and Communications Technologies) ecosystem now approaches 5% of world electricity consumption and this ICT energy use will continue to grow fast because of the information appetite of Big Data, large networks and large infrastructures as Clouds that unavoidably leads to large 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 infrastructures). 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 moderately loaded Cloud, 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 stakeholders. 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 convincing 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 their 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 support 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

- SimGrid was named as one of ten “French scientific successes in year 2018” in the French government report “Vers une loi de programmation pluriannuelle de la Recherche” [0]
- The FogGuru European project has started a real-life experimentation in València (Spain) of Fog computing technologies applied to smart water supply management, in collaboration with Emivasa, the public-private company in charge of water supply.
- The RI/RE project (funded by the CNRS Momentum call) has started in 2019. This project will strengthen our exploration of Smart Grids relations with computing systems.

6. New Software and Platforms

6.1. 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/

6.2. 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

[0] “Vers une loi de programmation pluriannuelle de la Recherche.” French government’s press release, Feb 2019, page 6. https://cache.media.enseignementsup-recherche.gouv.fr/file/Recherche/91/7/dp-loi_programmation_1069917.pdf

6.3. 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.

NEWS OF THE YEAR: There were 3 major releases in 2019: Python bindings were introduced, SMPI now partially supports some of the MPI/IO functions, a new model for Wifi networks was proposed, and the API for the simulation of storage resources was completely revisited. We also pursued our efforts to improve the documentation of the software, simplified the web site, and made a lot of bug fixing and code refactoring.

- Participants: Adrien Lèbre, Arnaud Legrand, Augustin Degomme, Florence Perronnin, Frédéric Suter, Jean-Marc Vincent, Jonathan Pastor, Luka Stanisic and Martin Quinson
- Partners: CNRS - ENS Rennes
- Contact: Martin Quinson
- URL: https://simgrid.org/

6.4. 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.

- Authors: Yvon Jégou, Manh Linh Pham, Nikolaos Parlavantzas and Christine Morin
- Contact: Nikolaos Parlavantzas
- Publication: hal-01612979/
- URL: https://team.inria.fr/myriads/software-and-platforms/diffuse/

6.5. GinFlow

KEYWORDS: Workflow - Distributed computing - Distributed - Distributed Applications - Dynamic adaptation - 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

6.6. libcvss

KEYWORD: Cybersecurity

FUNCTIONAL DESCRIPTION: libcvss is a Rust implementation of the CVSS specification. The supported versions of CVSS are 2.0, 3.0 and 3.1.

The official CVSS website describes CVSS this way: "The Common Vulnerability Scoring System (CVSS) provides a way to capture the principal characteristics of a vulnerability and produce a numerical score reflecting its severity. The numerical score can then be translated into a qualitative representation (such as low, medium, high, and critical) to help organizations properly assess and prioritize their vulnerability management processes."

libcvss provides Rust users with a native way to manipulate CVSS-formatted vulnerability data. Rust is leveraged to provide a CVSS implementation focused on both performance and correctness.

- Participant: Clement El Baz
- Contact: Clement El Baz
- URL: https://crates.io/crates/libcvss

7. New Results

7.1. Scaling Clouds

7.1.1. Efficient Docker container deployment in fog environments

Participants: Arif Ahmed, Lorenzo Civolani, Guillaume Pierre, Paulo Rodrigues de Souza Junior.

Fog computing aims to extend datacenter-based cloud platforms with additional computing, 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 traditional clouds: to provide computing resources in the 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 ultra-high-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. Significant improvements in the Docker image cache hit rate can be obtained by sharing the caches of multiple co-located servers rather than letting them operate independently [9]. In the case when an image must be downloaded and locally installed, large performance gains can be obtained with relatively simple modifications in the way Docker imports container images [3]. Finally, we showed (in collaboration with Prof. Paolo Bellavista from the University of Bologna) that it is possible to let a container start producing useful work even before its image has been fully downloaded [14]. Another paper in this direction of work is in preparation about the way to speedup the boot phase of Docker containers. We are also exploring innovative techniques to improve the performance of live container migration in fog computing environments.

7.1.2. Fog computing platform design

Participants: Ali Fahs, Ayan Mondal, Nikos Parlavantzas, Guillaume Pierre, Mulugeta Tamiru.

There does not yet exist any reference platform for fog computing platforms. We therefore investigated how Kubernetes could be adapted to support the specific needs of fog computing platforms. In particular we focused on the problem of redirecting end-user traffic to a nearby instance of the application. When different users impose various load on the system, any traffic routing system must necessarily implement a tradeoff between proximity and fair load-balancing between the application instances. We demonstrated how such customizeable traffic routing policies can be integrated in Kubernetes to help transform it in a suitable platform for fog computing [15]. We extended this work to let the platform automatically choose (and maintain over time) the best locations where application replicas should be deployed. A paper on this topic is currently under submission. We finally started addressing the topic of application autoscaling such that the system can enforce performance guarantees despite traffic variations. We expect one or two publications on this topic next year.

In collaboration with Prof. Misra from IIT Kharagpur (India), and thanks to the collaboration established by the FogCity associate team, we developed mechanisms based on game theory to assign resources to competing applications in a fog computing platform. The objective of those mechanisms is to satisfy user preferences while maximizing resource utilisation. We evaluated the mechanisms using an emulated fog platform built on Kubernetes and Grid'5000, and showed that they significantly outperform baseline algorithms. A paper on this topic is in preparation.

7.1.3. Edgification of micro-service applications

Participants: Genc Tato, Cédric Tedeschi, Marin Bertier.

Last year, we investigated in collaboration with Etienne Riviere from UC Louvain the feasibility and possible benefits brought about by the *edgification* of a legacy micro-service-based application [35]. In other words, we devised a method to classify services composing the application as *edgifiable* or not, based on several criteria. We applied this method to the particular case of the ShareLatex application which enables the collaborative edition of LaTeX documents. Recently, we continue this work by automate the localization and the migration of microservices. Our middleware, based on Koala [36], a lightweight Distributed Hash Table, allows adapting compatible legacy microservices applications for hybrid core/edge deployments [21].

7.1.4. Community Clouds

Participants: Jean-Louis Pazat, Bruno Stevant.

Small communities of people who need to share data and applications can now buy inexpensive devices in order to use only "on premise" resources instead of public Clouds. This "self-hosting-and-sharing" solution provides a better privacy and does not need people to pay any monthly fee to a resource provider. We have implemented a prototype based on micro-services in order to be able to distribute the load of applications among devices.

However, such a distributed platform needs to rely on a very good distribution of the computing and communication load over the devices. Using an emulator of the system, we have shown that, thanks to well known optimization techniques (Particle Swarm Optimization), it is possible to quickly find a service placement resulting in a response time close to the optimal one.

This year we evaluated the results of the optimization algorithm on a prototype (5 "boxes" installed in different home locations connected by fiber or ADSL). Results shown that due to the variation of the network available bandwidth it is necessary to dynamically modify the deployment of applications. This was not a big surprise, but we were not able to find any predictive model of this variation during a day. So, we developed and experimented a dynamic adaptation of the placement of micro-services based applications based on a regular monitoring of the response time of applications. We plan to submit a paper on this topic in early 2020.

7.1.5. Geo-distributed data stream processing

Participants: Hamidreza Arkian, Davaadorj Battulga, Mehdi Belkhiria, Guillaume Pierre, Cédric Tedeschi.

We investigated a decentralized scaling mechanism for stream processing applications where the different operators composing the processing topology are able to take their own scaling decisions independently, based on local information. We built a simulation tool to validate the ability of our algorithm to react to load variation. Then, we started the development of a software prototype of a decentralized Stream Processing Engine including this autoscaling mechanism, and deployed it over the Grid'5000 platform. Two papers have been accepted in 2019 about this work [11], [12].

Although data stream processing platforms such as Apache Flink are widely recognized as an interesting paradigm to process IoT data in fog computing platforms, the existing performance model to capture of stream processing in geo-distributed environments are theoretical works only, and have not been validated against empirical measurements. We developed and experimentally validated such a model to represent the performance of a single stream processing operator [10]. This model is very accurate with predictions $\pm 2\%$ of the actual values even in the presence of heterogeneous network latencies. Individual operator models can be composed together and, after the initial calibration of a first operator, a reasonably accurate model for other operators can be derived from a single measurement only.

7.1.6. QoS-aware and energy-efficient resource management for Function-as-a-Service

Participants: Yasmina Bouizem, Christine Morin, Nikos Parlavantzas.

Recent years have seen the widespread adoption of serverless computing, and in particular, Function-as-a-Service (FaaS) systems. These systems enable users to execute arbitrary functions without managing underlying servers. However, existing FaaS frameworks provide no quality of service guarantees to FaaS users in terms of performance and availability. Moreover, they provide no support for FaaS providers to reduce energy consumption. The goal of this work is to develop an automated resource management solution for FaaS plaforms that takes into account performance, availability, and energy efficiency in a coordinated manner. This work is performed in the context of the thesis of Yasmina Bouizem. In 2019, we integrated a fault-tolerance mechanism into Fission, an open-source FaaS framework based on Kubernetes, and are currently evaluating its impact on performance, availability, and energy consumption.

7.2. Greening Clouds

7.2.1. Energy Models

Participants: 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 infrastructures). 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. These results have been published in [24].

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 should be based on experimental measurement campaigns on heterogeneous devices. As hardware architectures become more complex, measurement campains are required to better understand their energy consumption and to identify potential sources of energy waste. These results, conducted with Amina Guermouche (IMT Telecom SudParis), have been presented in [30].

Similarly, software stacks add complexity in the identification of energy inefficiencies. For HPC applications, precise measurements are required to determine the most efficient options for the runtime, the resolution algorithm and the mapping on physical resources. An example of such a study has been published in collaboration with HiePACS (Bordeaux) and NACHOS (Sophia) teams in [8].

The fine-grain measurements lead us to propose models that have been used to compare different Cloud architectures (from fog and edge to centralized clouds) in terms of energy consumption on a given scenario. These results have been published in [4].

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. In particular, this year we implemented in SimGrid a flow-based energy model for wired network devices [17].

7.2.2. End-to-end energy models for the Internet of Things

Participants: Anne-Cécile Orgerie, Loic Guegan.

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 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 Cloud, depending on the number of devices and the desired application QoS. This work has been published in [18].

7.2.3. Exploiting renewable energy in distributed clouds

Participants: Benjamin Camus, Anne-Cécile Orgerie.

The growing appetite of Internet services for Cloud resources leads to a consequent increase in data center (DC) facilities worldwide. This increase directly impacts the electricity bill of Cloud providers. Indeed, electricity is currently the largest part of the operation cost of a DC. Resource over-provisioning, energy non-proportional behavior of today's servers, and inefficient cooling systems have been identified as major contributors to the high energy consumption in DCs.

In a distributed Cloud environment, on-site renewable energy production and geographical energy-aware load balancing of virtual machines allocation can be associated to lower the brown (i.e. not renewable) energy consumption of DCs. Yet, combining these two approaches remains challenging in current distributed Clouds. Indeed, the variable and/or intermittent behavior of most renewable sources – like solar power for instance – is not correlated with the Cloud energy consumption, that depends on physical infrastructure characteristics and fluctuating unpredictable workloads.

7.2.4. Smart Grids

Participants: Anne Blavette, Benjamin Camus, Anne-Cécile Orgerie, Martin Quinson.

Smart grids allow to efficiently perform demand-side management in electrical grids in order to increase the integration of fluctuating and/or intermittent renewable energy sources in the energy mix. In this work, we consider the computing infrastructure that controls the smart grid. This infrastructure comprises communication and computing resources to allow for a smart management of the electrical grid. In particular, we study the influence of communication latency over a shedding scenario on a small-scale electrical network. We show that depending on the latency some shedding strategies are not feasible [13].

7.3. Securing Clouds

7.3.1. Security monitoring in Cloud computing platforms

Participants: Clément Elbaz, Christine Morin, Louis Rilling, Amir Teshome Wonjiga.

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 the past years we proposed a verification method for security monitoring SLOs [37] and we have then studied a methodology to define security monitoring SLOs that are at the same time relevant for the tenant, achievable for the provider, and verifiable. The methodology is based on metrics benchmarks that a cloud service provider runs on a set of basic setups of an NIDS (Network Intrusion Detection), the basic setups covering together the variety of NIDS rules that may interest tenants. In order to make it achievable for a cloud service provider to run such benchmarks despite thousands of rules that could be chosen individually by tenants, we proposed a rule clustering strategy to lower the number of sets of rules that should be benchmarked and thus the number of benchmarks run. Finally we proposed extensions to an existing cloud SLA language to define security monitoring SLOs. These results were published in a technical report [27] as well as in Amir Teshome Wonjiga's thesis (to appear) and were submitted for publication in an international conference.

In a side project with Dr Sean Peisert at LBNL, the work on security SLO verification was extended to the use case of data integrity, where tenants outsource data to a cloud storage provider. This work allowed us to tackle a challenge in SLO verification because, in this use case as well as in the security monitoring use case, tenants cannot verify SLOs without a minimal trust in providers involvment in the verification process. We proposed a strategy based on blockchains that allows tenants as well as providers to do SLO verification without having to trust any individual entity. This work was published in the CIFS security workshop [22].

To make security monitoring SLOs adaptable to context changes like the evolution of threats and updates to the tenants' software, we have worked on automating the mitigation of new threats during the time window in which no intrusion detection rule exist and no security patch is applied yet (if available). This time winwow is critical because newly published vulnerabilities get exploited up to five orders of magnitude right after they are published and the time window may last several days or weeks. We have worked on a first step of mitigation, which consists in deciding if a newly published vulnerabiliy impacts a given information system. A major challenge in automating this step is that newly published vulnerabilities do not contain machine-readable data and this data only appears up to several weeks later. For this reason we designed and evaluated a keyword extraction process from the free-form text description of a vulnerability to map a given vulnerability to product names. This keyword exctraction process was first published at the RESSI French security conference [23] and will appear in the NOMS 2020 international conference. In future work this mapping should be combined with a knowledge base of the information system to automatically score the impact of a new vulnerability on the information system.

Our results were published in [27], [28], [22], [23], [25].

7.3.2. Privacy monitoring in Fog computing platforms

Participants: Mozhdeh Farhadi, Guillaume Pierre.

IoT devices are integrated in our daily lives, and as a result they often have access to lots of private information. For example many digital assistants (Alexa, Amazon Echo...) were shown to have violated the privacy policy they had established themselves. To increase the level of confidence that end users may have in these devices and the applications which process their data, we started designing monitoring mechanisms such that the fog or the cloud platform can certify whether an application actually follows its own privacy policy or not. A survey paper on security of fog computing platforms is under submission, and we expect another paper on privacy monitoring in 2020.

7.4. Experimenting with Clouds

7.4.1. Simulating distributed IT systems

Participants: Toufik Boubehziz, Benjamin Camus, Anne-Cécile Orgerie, Millian Poquet, Martin Quinson.

Our team plays a major role in the advance of the SimGrid simulator of IT systems. This framework has a major impact on the community. Cited by over 900 papers, it was used as a scientific instrument by more than 300 publications over the years.

This year, we pursued our effort to ensure that SimGrid becomes a *de facto* standard for the simulation of distributed IT platforms. We further polished the new interface to ensure that it correctly captures the concepts needed by the experimenters, and provided a Python binding to smooth the learning curve. To that extend, we also continued our rewriting of the documentation.

The work on SimGrid is fully integrated to the other research efforts of the Myriads team. This year, we added the ability to co-simulate IT systems with SimGrid and physical systems modeled with equational systems [13]. This work, developed to study the co-evolution of thermal systems or of the electic grid with the IT system, is now distributed as an official plugin of the SimGrid framework.

7.4.2. Formal methods for IT systems

Participants: Ehsan Azimi, The Anh Pham, Martin Quinson.

The SimGrid framework also provide a state of the art Model-Checker for MPI applications. This can be used to formally verify whether the application entails synchronization issues such as deadlocks or livelocks [32]. This year, we pursued our effort on this topic, in collaboration with Thierry Jéron (EPI SUMO).

The Anh Pham defended his thesis this year on techniques to mitigate the state space explosion while verifying asynchronous distributed applications. He adapted an algorithm leveraging event folding structures to this context. This allows to efficiently compute how to not explore equivalent execution traces more than once. This work was published this year[19]. This work, co-advised by Martin Quinson with Thierry Jéron (team SUMO, formal methods), was important to bridge the gap between the involved communities.

Ehsan Azimi joined the Myriads team as an engineer in December to integrate the results of this thesis into the SimGrid framework.

7.4.3. Executing epidemic simulation applications in the Cloud

Participants: Christine Morin, Nikos Parlavantzas, Manh Linh Pham.

In the context of the DiFFuSE ADT and in collaboration with INRA researchers, we transformed a legacy application for simulating the spread of Mycobacterium avium subsp. paratuberculosis (MAP) to a cloud-enabled application based on the DiFFuSE framework (Distributed framework for cloud-based epidemic simulations). This is the second application to which the DiFFuSE framework is applied. The first application was a simulator of the spread of the bovine viral diarrhea virus, developed within the MIHMES project (2012-2017). Using both the MAP and BVDV applications, we performed extensive experiments showing the advantages of the DiFFuSE framework. Specifically, we showed that DiFFuSE enhances application performance and allows exploring different cost-performance trade-offs while supporting automatic failure handling and elastic resource acquisition from multiple clouds [7].

7.4.4. Tools for experimentation

Participant: Matthieu Simonin.

In collaboration with the STACK team and in the context of the Discovery IPL, novel experimentation tools have been developed. In this context experimenting with large software stacks (OpenStack, Kubernetes) was required. These stacks are often tedious to handle. However, practitioners need a right abstraction level to express the moving nature of experimental targets. This includes being able to easily change the experimental conditions (e.g underlying hardware and network) but also the software configuration of the targeted system (e.g service placement, fined-grained configuration tuning) and the scale of the experiment (e.g migrate the experiment from one small testbed to another bigger testbed).

In this spirit we discuss in [31] a possible solution to the above desiderata. We illustrate its use in a real world use case study which has been completed in [34]. We show that an experimenter can express their experimental workflow and execute it in a safe manner (side effects are controlled) which increases the repeatability of the experiments.

The outcome is a library (EnOSlib) target reusability in experiment driven research in distributed systems. The library can be found in https://bil.inria.fr/fr/software/view/3589/tab.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. INDIC - Cybersecurity Pole of Excellence (2014-2020)

Participants: Clément Elbaz, Christine Morin, Louis Rilling, Amir Teshome Wonjiga.

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 funded the PhD of Anna Giannakou (defended in 2017) and Amir Teshome Wonjiga (defended in 2019). Clément Elbaz is partially funded by the Brittany Regional Council in the PEC framework.

8.2. National Initiatives

8.2.1. ADEME RennesGrid (2017-2020)

Participants: Anne Blavette, 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 closely collaborate with Anne Blavette (CR CNRS in electrical engineering, SATIE, Rennes) and co-supervise the post-doc of Benjamin Camus who started in April 2018 on evaluating the impact of the IT infrastructure in the management of smart grids.

8.2.2. Inria ADT Mc SimGrid (2019-2021)

Participants: Ehsan Azimi, Martin Quinson.

The Mc SimGrid technological development action funded by INRIA targets the refactoring of model checker that is integrated to the SimGrid simulation framework. Its software quality should be improved to be on par with the rest of the SimGrid framework. Our ultimate goal is to make this model-checker usable in production, both to assess real-size applications and as a workbench for the researchers designing new techniques and algorithms for the verification of distributed asynchronous applications and algorithms.

The technical actions envisioned for this ADT are the complete re-factoring of this software module, and the exposure of a sensible python interface to experiment with new exploration algorithms. This work is lead by Ehsan Azimi, in collaboration with Thierry Jéron from the Sumo team.

8.2.3. Inria IPL Discovery (2015-2019)

Participants: 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.

8.2.4. Inria IPL Hac Specis (2016-2020)

Participants: Dorra Boughzala, Anne-Cécile Orgerie, The Anh Pham, Martin Quinson.

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.

The Anh Pham defended his thesis on December 6., on techniques to mitigate the state space explosion while verifying asynchronous distributed applications. He proposed a new algorithm to mitigate the state space explosion problem (published this year [19]), using event folding structures to efficiently compute how to not explore equivalent execution traces more than once. This work, co-advised by Martin Quinson with Thierry Jéron (team SUMO, formal methods), was important to bridge the gap between the involved communities.

During her PhD thesis, Dorra Boughzala studied the energy consumption of GPU and the simulation tools of the literature related to this aspect. Her work is co-advised by Laurent Lefèvre (Avalon team, Lyon), Martin Quinson and Anne-Cécile Orgerie.

8.2.5. SESAME ASTRID project (2016-2019)

Participants: Mehdi Belkhiria, Pascal Morillon, Christine Morin, Matthieu Simonin, Cédric Tedeschi.

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. In 2019, we focused on autoscaling and placement for Stream Processing applications.

In 2019, we investigated the dynamic, decentralized scaling of stream processing applications. Also, we collaborated with the Inria OBELIX team to scale and deploy a machine learning application they developed to build a model of a *normal* vessel trajectory.

8.2.6. 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.

8.3. European Initiatives

8.3.1. H2020 Projects

8.3.1.1. H2020 MSCA FogGuru

Participants: Hamidreza Arkian, Davaadorj Battulga, Mozhdeh Farhadi, Julie Montégu, Guillaume Pierre, Mulugeta Ayalew Tamiru, Cédric Tedeschi, Paulo Rodrigues de Souza Junior.

Title: FogGuru – Training the Next Generation of European Fog Computing Experts

Program: H2020 MSCA ITN EID

Duration: September 2017 - August 2021

Coordinator: Guillaume Pierre

Participants:

- University of Rennes 1, France (coordinator)
- Technische Universität Berlin, Germany
- Elastisys AB, Sweden
- U-Hopper srl, Italy
- EIT Digital Rennes, France
- Las Naves, Spain

Abstract: 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 innovation process stemming from fundamental research towards invention and development of innovative products and services, and to real-life deployment, experimentation and engagement with beta-testers.

8.3.2. Collaborations in European Programs, Except FP7 & H2020

8.3.2.1. EIT Digital DriveTrust

Participant: Guillaume Pierre.

Program: EIT Digital

Project acronym: DriveTrust

Project title: AI-Powered Driving Evaluation

Duration: January 2019 - December 2019

Coordinator: University of Rennes 1, France

Other partners:

Eurapco, Switzerland

Achmea, the Netherlands

Imec, Belgium

Abstract: This project aims to develop and commercialize an AI-powered dash cam with short range V2X and LTE communication capabilities. The product uses the newest AI capable hardware for real-time object detection. The device can detect street signs, traffic lights, other cars, and pedestrians. Combined with sensor data from the accelerometer, GPS and weather data from the cloud we use the data to calculate different dimensions of driving profiles. In addition the V2X and object detection capabilities allow us to warn the driver in real-time about dangers on the road.

8.3.3. Inria Associate Teams Not Involved in an Inria International Labs

8.3.3.1. FogCity

Participants: Ayan Mondal, Nikos Parlavatzas, Guillaume Pierre.

Title: QoS-aware Resource Management for Smart Cities

International Partner (Institution - Laboratory - Researcher):

IIT Kharagpur (India) - Department of Computer Science and Engineering - Sudip Misra

IIT Kanpur (India) - Department of Industrial and Management Engineering - Subhas Chandra Misra

Start year: 2018

See also: https://team.inria.fr/myriads/projects/fogcity/

Abstract: The FogCity associate team proposal concerns a collaboration between the Myriads project-team, and two research teams at Indian Institute of Technology Kharagpur and Indian Institute of Technology Kanpur. The proposal focuses on a smart city scenario in which data from static and mobile sensors is routed to appropriate fog data centres based on application QoS requirements. The main goal of the research is to select suitable nodes within the fog data centers to optimize the QoS of the applications in terms of latency. The different teams have complementary expertise in theoretical research (Indian partners) and system research (Inria Myriads project-team) and share a strong research interest in IoT and Fog Computing.

8.3.3.2. FogRein

Participants: Anne-Cécile Orgerie, Martin Quinson.

Title: Steering Efficiency for Distributed Applications

International Partner: Gene Cooperman, College of Computer and Information Science, Northeastern University (USA).

Start year: 2019

In Fog Computing, the Internet of Things (IoT), and Intermittent Computing, low-power devices migrate geographically, and are required to rapidly assimilate new data in a power-efficient manner. This is a key component of any Smart Interfaces solution as devices migrate from the IT infrastructure to the Edge of the Cloud in order to provide Function-as-a-Service, High-availability mobility, and IT infrastructure malleability. A three-tier strategy is proposed toward steering Fog applications in order to optimize the energy efficiency and sustainability. The strategy will leverage the backgrounds of the participants in Fog Computing, checkpointing, scheduling, Green Levers within the IT infrastructure, and a simulation infrastructure for predicting and efficiently steering such distributed applications. The Inria team and the Northeastern team are uniquely positioned to make rapid progress due to their long history of collaborative research based on visits by both permanent members and PhD students in the two directions.

8.3.4. Inria International Partners

8.3.4.1. Informal International Partners

UC Louvain (Belgium): We collaborate with Prof. Etienne Riviere on legacy application edgification. Genc Tato spent six month at UCL.

Tlemcen University (Algeria): We collaborate with Dr. Djawida Dib on energy-efficient and fault-tolerant resource management in containerized clouds. Christine Morin and Nikos Parlavantzas are co-advising Yasmina Bouizem, who is enrolled in both Tlemcen University and University of Rennes 1.

University of Bologna (Italy): We collaborate with Prof. Paolo Bellavista on the design of performance-efficient fog computing platforms. Lorenzo Civolani from University of Bologna spent 6 months in the Myriads team to complete his master thesis internship. A paper on his work has been accepted for publication [14].

Umeå University (Sweden): We collaborate with Prof. Erik Elmroth on the control of large-scale cloud and fog computing platforms. Ali Fahs spent 6 months at Umeå University where he worked on autoscaling techniques for future fog computing platforms.

Rutgers University (USA): We collaborate with Prof. Manish Parashar on improving the energy efficiency of distributed data-intensive applications. This collaboration is outlined by a joint publication in [20].

University of Hawaii (USA): We collaborate with Prof. Henri Casanova on simulating the energy consumption of scientific workflows. This collaboration is outlined by a joint publication in [16].

University of Southern California (USA): We collaborate with Dr. Rafael Ferreira da Silva and Prof. Ewa Deelman on simulating the energy consumption of scientific workflows. This collaboration is outlined by a joint publication in [16].

8.3.5. Participation in Other International Programs

8.3.5.1. Inria International Chairs

Deborah AGARWAL

Title: Workflow, user centered design, and data management as well as mobile applications for data science

International Partner (Institution - Laboratory - Researcher):

Université californienne de Santa Barbara (United States) - Computational Research Division - Deborah Agarwal

Duration: 2015 - 2019

Start year: 2015

8.4. International Research Visitors

8.4.1. Visits of International Scientists

8.4.1.1. Internships

Lorenzo CIVOLANI

Date: Sep 2018 - Feb 2019

Institution: University of Bologna (Italy)

Supervisor: Guillaume Pierre

Adrien GOUGEON

Date: Feb 2019 - Jun 2019

Institution: ENS Rennes

Supervisors: Anne-Cécile Orgerie and Benjamin Camus

Archana WALE

Date: Jan 2019 - Jun 2019

Institution: University of Rennes 1

Supervisor: Guillaume Pierre

Romain Olivo

Date: Juin 2019 - Aug 2019

Institution: Inria

Supervisor: Matthieu Simonin

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Member of the Organizing Committees

- Guillaume Pierre is a member of the ACM/IFIP Middleware steering committee.

9.1.2. Scientific Events: Selection

9.1.2.1. Chair of Conference Program Committees

- Guillaume Pierre is co-chair of the Intelligent Systems and Infrastructure track of the TheWebConf 2020 conference.
- Anne-Cécile Orgerie is co-chair of the Architecture and Networking track of the IEEE/ACM CCGrid 2019 conference.
- Anne-Cécile Orgerie is global co-chair of the Cluster and Cloud Computing track of the EuroPar 2019 conference.

9.1.2.2. Member of the Conference Program Committees

- Guillaume Pierre was a PC member of the TheWebConf 2019 conference.
- Guillaume Pierre was a PC member of the ACM/IEEE CCGRID 2019 conference.
- Guillaume Pierre was a PC member of the ACM/IFIP Middleware 2019 conference.
- Guillaume Pierre was a PC member of the Conpas 2019 conference.

- Guillaume Pierre was a PC member of the InterCloud-HPC 2019 workshop.
- Mozhdeh Farhadi was a PC member of the CIFS 2019 workshop.
- Anne-Cécile Orgerie was a PC member of the IEEE HiPC 2019 conference.
- Anne-Cécile Orgerie was a PC member of the ICA3PP 2019 conference.
- Anne-Cécile Orgerie was a PC member of the IEEE/ACM CCGrid 2019 conference.
- Anne-Cécile Orgerie was a PC member of the ICPP 2019 conference.
- Anne-Cécile Orgerie was a PC member of the AlgoTel 2019 conference.
- Anne-Cécile Orgerie was a PC member of the IEEE IC2E 2019 conference.
- Cédric Tedeschi was a PC member of the ICPP 2019 conference.
- Cédric Tedeschi was a PC member of the IEEE ICWS 2019 conference.
- Cédric Tedeschi was a PC member of the Conpas 2019 conference.
- Nikos Parlavantzas was a PC member of the UCC 2019 conference.
- Nikos Parlavantzas was a PC member of the CloudCom 2019 conference.
- Nikos Parlavantzas was a PC member of the ISPDC 2019 conference.
- Nikos Parlavantzas was a PC member of the CrossCloud 2019 workshop.
- Nikos Parlavantzas was a PC member of the VHPC 2019 workshop.

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

- Anne-Cécile Orgerie is member of the editorial board of International Journal of Distributed Sensor Networks, SAGE Publishing.

9.1.3.2. Reviewer - Reviewing Activities

- Guillaume Pierre reviewed one article for the IEEE Networking Letters.
- Guillaume Pierre reviewed one article for the Software: Practice and Experience journal.
- Cédric Tedeschi reviewed one article for the TPDS journal.
- Nikos Parlavantzas reviewed one article for the Applied Computing and Informatics journal.
- Marin Bertier reviewed one article for the Journal of Supercomuting.
- Martin Quinson reviewed one article for the Journal of Computational Design and Engineering.

9.1.4. Invited Talks

- Guillaume Pierre gave a keynote speech at the IEEE CloudNet 2019 conference: “From Cloud to Fog: The Tao of IT Infrastructure Decentralization.”
- Anne-Cécile Orgerie gave a seminar at LORIA lab in September 2019.
- Anne-Cécile Orgerie gave a seminar at the scientific council of INS2I in September 2019.
- Anne-Cécile Orgerie gave a talk at the Cloud Control Workshop in June 2019.
- Martin Quinson gave a talk at the "Algorithmes et programmation" workshop, at Centre International des Rencontres Mathématiques (CIRM) in May 2019.

9.1.5. Leadership within the Scientific Community

- Anne-Cécile Orgerie is vice-chair of ASF, the French Chapter of ACM SIGOPS.

9.1.6. Scientific Expertise

- 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.

9.1.7. Research Administration

- Anne-Cécile Orgerie is an 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.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Bachelor: Marin Bertier, Networks, Département Informatique L3, Insa Rennes.

Bachelor: Marin Bertier, Language C, Département Informatique L3, Insa Rennes.

Bachelor: Marin Bertier, Language C, Département Mathématique L3, Insa Rennes.

Bachelor: Jean-Louis Pazat, Introduction à la programmation, L1, Département STPI, INSA de Rennes.

Bachelor: Jean-Louis Pazat, Calcul Hautes Performances, Département Informatique L3, Insa Rennes.

Bachelor: Jean-Louis Pazat, Calcul Hautes Performances, Département Mathematiques L3, Insa Rennes.

Bachelor: Guillaume Pierre, Systèmes Informatiques, L3 MIAGE, Univ. Rennes 1.

Bachelor: Guillaume Pierre, Systèmes d’exploitation, L3 Informatique, Univ. Rennes 1.

Bachelor: Martin Quinson, Architecture et Systèmes, 60 hETD, L3 Informatique, ENS Rennes.

Bachelor: Martin Quinson, Pedagogy, 15 hETD, L3 Informatique, ENS Rennes.

Master: Marin Bertier, Operating Systems, Département Informatique M1, INSA de Rennes

Master: Marin Bertier, Distributed systems, Département Informatique M2, INSA de Rennes

Master: Anne-Cécile Orgerie, Cloud & Big Data, 30 hETD, M1, ENS Rennes.

Master: Anne-Cécile Orgerie, Green ICT, 4.5 hETD, M2, Telecom SudParis Evry.

Master: Anne-Cécile Orgerie, Green ICT,3 hETD, M2, ENSSAT Lannion.

Master: Anne-Cécile Orgerie, Green IT, 6 hETD, M1, INSA Rennes.

Master: Nikos Parlavantzas, Clouds, M1, INSA Rennes.

Master: Nikos Parlavantzas, Performance Evaluation, M1, INSA Rennes.

Master: Nikos Parlavantzas, Project in Large-Scale Systems, M2, INSA Rennes.

Master: Nikos Parlavantzas, Big Data and Applications, M1, INSA Rennes.

Master: Jean-Louis Pazat, Parallel Computing, M1 Département Informatique Insa Rennes.

Master: Jean-Louis Pazat, Internet Of Things, M1 & M2 Département Informatique Insa Rennes.

Master: Guillaume Pierre, Distributed Systems, M1, Univ. Rennes 1.

Master: Guillaume Pierre, Service technology, M1, Univ. Rennes 1.

Master: Guillaume Pierre, Techniques de développement pour le cloud, M2, Univ. Rennes 1.

Master: Guillaume Pierre, Advanced Cloud Infrastructures, M2, Univ. Rennes 1.

Master: Martin Quinson, Préparation à l’Agrégation de Science Industrielle (Programming and Software Engineering, 20h ETD; Operating Systems and C programming, 20 hETD; Networking, 20h ETD), ENS Rennes.

Master: Martin Quinson, Scientific Outreach, M2, 30 hEDT, ENS Rennes.

Master: Cédric Tedeschi, Concurrency in Systems and Networks, M1, Univ. Rennes 1.

Master: Cédric Tedeschi, Service Technology, M1, Univ. Rennes 1.

Master: Cédric Tedeschi, Parallel Programming, M1, Univ. Rennes 1.

Master: Cédric Tedeschi, Systems, L3, Univ. Rennes 1.

9.2.2. Supervision

Defended PhD: Genc Tato, "Locality-aware Lazy Overlay Networks for WANs," defended on December 19th 2019, supervised by Marin Bertier, Cédric Tedeschi.

Defended PhD: The Anh Pham, "Dynamic Formal Verification of High Performance Runtimes and Applications," defended on December 6th 2019, supervised by Martin Quinson, Thierry Jéron.

Defended PhD: Amir Teshome Wonjiga, "User-Centric Security Monitoring in Cloud Environments", defended on June 3rd 2019, supervised by Louis Rilling and Christine Morin.

PhD in progress: Adrien Gougeon, "Designing an energy-efficient communication network for the dynamic and distributed control of the electrical grid", started in September 2019, supervised by Anne-Cécile Orgerie and Martin Quinson.

PhD in progress: Paulo Rodrigues De Souza Junior, "fog computing service roaming techniques", started in December 2018, supervised by Guillaume Pierre and Daniele Miorandi (U-Hopper srl, Italy).

PhD in progress: Davaadorj Battulga, "Scalable data pipelines for fog computing applications", started in September 2018, supervised by Cédric Tedeschi and Daniele Miorandi (U-Hopper srl, Italy).

PhD in progress: Mulugeta Tamiru, "Automatic optimization of autonomous management systems", started in September 2018, supervised by Guillaume Pierre and Erik Elmroth (Elastisys AB, Sweden).

PhD in progress: Mozhdeh Farhadi, "Fog computing-enabled IoT situation-aware services", started in June 2018, supervised by Guillaume Pierre and Daniele Miorandi (U-Hopper srl, Italy).

PhD in progress: Hamidreza Arkian, "Stream processing operator placement", started in June 2018, supervised by Guillaume Pierre and Erik Elmroth (Elastisys AB, Sweden).

PhD in progress: Mehdi Belkhiria, "Dynamic Stream Processing for Maritime Traffic Surveillance," started in December 2017, supervised by Cédric Tedeschi.

PhD in progress (*co-tutelle*): Yasmina Bouizem, "Energy-efficient, fault-tolerance mechanisms for containerized cloud applications," started in December 2017, supervised by Didi Fedoua (Tlemcen University, Algeria), Djawida Dib (Tlemcen University, Algeria), Christine Morin and Nikos Parlavantzas.

PhD in progress: Ali Jawad Fahs, "Decentralized Fog Computing Infrastructure Control", started in October 2017, supervised by Guillaume Pierre.

PhD in progress: Clément El Baz, "Reactive security monitoring in clouds," started in October 2017, supervised by Louis Rilling and Christine Morin.

PhD in progress: Loic Guegan, "Simulating Internet of Things", started in October 2017, supervised by Martin Quinson and Anne-Cécile Orgerie.

PhD in progress: Dorra Boughzala, "Simulating Energy Consumption of Continuum Computing between Heterogeneous Numerical Infrastructures in HPC", started in December 2017, supervised by Laurent Lefèvre (Avalon team in Lyon), Anne-Cécile Orgerie and Martin Quinson.

PhD in progress: Arif Ahmed, "Scalable Decentralized Edge Cloud Infrastructures", started in October 2016, supervised by Guillaume Pierre.

PhD in progress: Baptiste Goupille-Lescar, "Designing agile, distributed cyber-physical systems with advanced collaboration capabilities", started in January 2016, supervised by Eric Lenormand (Thales), Christine Morin, Nikos Parlavantzas.

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), supervised by Jean-Louis Pazat.

9.2.3. Juries

- Guillaume Pierre chaired the PhD committee of Alexandre Veith, ENS Lyon, September 23rd 2019.
- Guillaume Pierre chaired the PhD committee of Jad Darrous, ENS Lyon, December 17th 2019.
- Martin Quinson chaired the PhD committee of Pierre Huchant, Ecole Nationale Supérieure d'Électronique, Informatique, Télécommunications, Mathématique et Mécanique de Bordeaux (ENSEIRB-MATMECA), March 29th 2019.
- Anne-Cécile Orgerie is a member of the jury for *Agrégation externe de Sciences Industrielles de l'Ingénieur option ingénierie informatique*.
- Martin Quinson is a member of the jury for *Capes d'informatique*.
- Anne-Cécile Orgerie was a member of the PhD committee of Ayham Kassab, Université de Franche-Comté, November 14th 2019.
- Anne-Cécile Orgerie was a member of the PhD committee of Gustavo Rostirolla, Université Toulouse 3, November 25th 2019.
- Louis Rilling was a member of the PhD committee of Guillaume Averlant, INSA Toulouse, October 2nd 2019.
- Louis Rilling was a guest of the PhD committee of Mohammad Mahdi Bazm, Université de Nantes, July 8th 2019.

9.3. Popularization

9.3.1. Internal or external Inria responsibilities

- Martin Quinson is on the scientific board of the Blaise Pascal foundation, boosting the scientific outreach in the domains of Maths and Computer Science.

9.3.2. Education

- Guillaume Pierre shared his experience as the coordinator of a H2020 Maria Skłodowska Curie Action project with CCRRDT, the Brittany Region's consultative committee for research and technological development. January 14th 2019.
- "L codent L créent" is an outreach program to send PhD students to teach Python to middle school students in 8 sessions of 45 minutes. Tassadit Bouadi (Lacodam), Camille Maumet (Empenn) and Anne-Cécile Orgerie (Myriads) are coordinating the local version of this program, initiated in Lille. The first session in Rennes occured in April 2019, and a new session has started for 2019-2020. The program is currently supported by: Fondation Blaise Pascal, ED MathSTIC, Université Rennes 1 and Fondation Rennes 1.
- "J'peux pas, j'ai informatique" is an outreach program to deconstruct IT stereotypes. Organized over one day, this program welcomes 110 college students (5ème) each year and is coordinated by Anne-Cécile Orgerie. The next event is planned in April 2020.
- Martin Quinson participated to the DIU EIL (Enseigner l'Informatique au Lycée), training existing maths and enginering teachers so that they can teach Informatics in their colleges.
- Martin Quinson is in charge of a M2 lecture on scientific vulgarization at ENS Rennes. The students are asked to come up with new activities (or improve existing ones) on their research field.
- Martin Quinson gave a doctoral teaching on December 13 on scientific outreach targeting either schools or science forums at IRISA laboratory.

9.3.3. Internal action

- Anne-Cécile Orgerie, Martin Quinson and Matthieu Simonin: organize experimenter meetings for the large-scale systems department of IRISA: https://xug.gitlabpages.inria.fr/meetings/

9.3.4. Creation of media or tools for science outreach

- Martin Quinson is working on a set of new outreach activities with his Master 2 class at ENS Rennes since several years. The most promising activities are then tested on field in schools during another lecture at ENS Rennes, at Licence 3 level. All available resources are freely available at https://github.com/InfoSansOrdi/pedago-rennes/.

10. Bibliography

Major publications by the team in recent years

[1] H. CASANOVA, A. LEGRAND, M. QUINSON, F. SUTER.*SMPI Courseware: Teaching Distributed-Memory Computing with MPI in Simulation*, in "EduHPC-18 - Workshop on Education for High-Performance Computing", Dallas, United States, November 2018, p. 1-10, https://hal.inria.fr/hal-01891513

Publications of the year

Doctoral Dissertations and Habilitation Theses

[2] T. A. PHAM.*Efficient state-space exploration for asynchronous distributed programs: Adapting unfolding-based dynamic partial order reduction to MPI programs*, ENS Rennes, December 2019, https://hal.archives-ouvertes.fr/tel-02420950

Articles in International Peer-Reviewed Journal

[3] A. AHMED, G. PIERRE.*Docker-pi: Docker Container Deployment in Fog Computing Infrastructures*, in "International Journal of Cloud Computing", 2019, p. 1-20, forthcoming, https://hal.inria.fr/hal-02271434

[4] E. AHVAR, A.-C. ORGERIE, A. LEBRE.*Estimating Energy Consumption of Cloud, Fog and Edge Computing Infrastructures*, in "IEEE Transactions on Sustainable Computing", April 2019, p. 1-12 [*DOI :* 10.1109/TSUSC.2019.2905900], https://hal.archives-ouvertes.fr/hal-02083080

[5] G. COOPERMAN, M. QUINSON.*Sthread: In-Vivo Model Checking of Multithreaded Programs*, in "The Art, Science, and Engineering of Programming", 2020, forthcoming, https://hal.inria.fr/hal-02449080

[6] M. FARHADI, J.-L. LANET, G. PIERRE, D. MIORANDI.*A systematic approach towards security in Fog computing: assets, vulnerabilities, possible countermeasures*, in "Software: Practice and Experience", 2020, forthcoming, https://hal.inria.fr/hal-02441639

[7] N. PARLAVANTZAS, L. M. PHAM, C. MORIN, S. ARNOUX, G. BEAUNÉE, L. QI, P. GONTIER, P. EZANNO.*A Service-based Framework for Building and Executing Epidemic Simulation Applications in the Cloud*, in "Concurrency and Computation: Practice and Experience", 2019, forthcoming, https://hal.archives-ouvertes.fr/hal-02312313

International Conferences with Proceedings

[8] E. AGULLO, L. GIRAUD, S. LANTERI, G. MARAIT, A.-C. ORGERIE, L. POIREL.*Energy Analysis of a Solver Stack for Frequency-Domain Electromagnetics*, in "PDP 2019 - 27th Euromicro International Conference on Parallel, Distributed and Network-Based Processing", Pavia, Italy, 2019 27th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), IEEE, February 2019, p. 385-391 [*DOI :* 10.1109/EMPDP.2019.8671555], https://hal.archives-ouvertes.fr/hal-02191331

[9] A. AHMED, G. PIERRE.*Docker Image Sharing in Distributed Fog Infrastructures*, in "CloudCom 2019 - 11th IEEE International Conference on Cloud Computing Technology and Science", Sydney, Australia, IEEE, December 2019, https://hal.inria.fr/hal-02304285

[10] H. ARKIAN, G. PIERRE, J. TORDSSON, E. ELMROTH.*An Experiment-Driven Performance Model of Stream Processing Operators in Fog Computing Environments*, in "SAC 2020 - ACM/SIGAPP Symposium On Applied Computing", Brno, Czech Republic, March 2020, p. 1-9, https://hal.inria.fr/hal-02394396

[11] M. BELKHIRIA, C. TEDESCHI.*A Fully Decentralized Autoscaling Algorithm for Stream Processing Applications*, in "Auto-DaSP 2019 - Third International Workshop on Autonomic Solutions for Parallel and Distributed Data Stream Processing", Göttingen, Germany, August 2019, p. 1-12, https://hal.inria.fr/hal-02171172

[12] M. BELKHIRIA, C. TEDESCHI.*Design and Evaluation of Decentralized Scaling Mechanisms for Stream Processing*, in "CloudCom 2019 - 11th IEEE International Conference on Cloud Computing Technology and Science", Sidney, Australia, IEEE, December 2019, https://hal.inria.fr/hal-02351108

[13] B. CAMUS, A. BLAVETTE, A.-C. ORGERIE, J.-B. BLANC-ROUCHOSSÉ.*Co-simulation of an electrical distribution network and its supervision communication network*, in "CCNC 2020 - IEEE Consumer Communications & Networking Conference", Las Vegas, United States, CCNC: IEEE Consumer Communications & Networking Conference, IEEE, January 2020, p. 1-6, https://hal.archives-ouvertes.fr/hal-02352832

[14] L. CIVOLANI, G. PIERRE, P. BELLAVISTA.*FogDocker: Start Container Now, Fetch Image Later*, in "UCC 2019 - 12th IEEE/ACM International Conference on Utility and Cloud Computing", Auckland, New Zealand, ACM, December 2019, p. 51-59 [*DOI :* 10.1145/3344341.3368811], https://hal.inria.fr/hal-02332679

[15] A. FAHS, G. PIERRE.*Proximity-Aware Traffic Routing in Distributed Fog Computing Platforms*, in "CCGrid 2019 - IEEE/ACM International Symposium in Cluster, Cloud, and Grid Computing", Larnaca, Cyprus, IEEE, May 2019, p. 1-10 [*DOI :* 10.1109/CCGRID.2019.00062], https://hal.inria.fr/hal-02048965

[16] R. FERREIRA DA SILVA, A.-C. ORGERIE, H. CASANOVA, R. TANAKA, E. DEELMAN, F. SUTER.*Accurately Simulating Energy Consumption of I/O-intensive Scientific Workflows*, in "ICCS 2019 - International Conference on Computational Science", Faro, Portugal, ICCS 2019 - International Conference on Computational Science, Springer, June 2019, p. 138-152 [*DOI :* 10.1007/978-3-030-22734-0_11], https://hal.archives-ouvertes.fr/hal-02112893

[17] L. GUEGAN, B. L. AMERSHO, A.-C. ORGERIE, M. QUINSON.*A Large-Scale Wired Network Energy Model for Flow-Level Simulations*, in "AINA 2019 - 33rd International Conference on Advanced Information Networking and Applications", Matsue, Japan, Advances in Intelligent Systems and Computing, Springer, March 2019, vol. 926, p. 1047-1058 [*DOI :* 10.1007/978-3-030-15032-7_88], https://hal.inria.fr/hal-02020045

[18] L. GUEGAN, A.-C. ORGERIE.*Estimating the end-to-end energy consumption of low-bandwidth IoT applications for WiFi devices*, in "CloudCom 2019 - 11th IEEE International Conference on Cloud Computing Technology and Science", Sydney, Australia, IEEE, December 2019, https://hal.archives-ouvertes.fr/hal-02352637

[19] T. A. PHAM, T. JÉRON, M. QUINSON.*Unfolding-based Dynamic Partial Order Reduction of Asynchronous Distributed Programs*, in "FORTE 2019 - 39th International Conference on Formal Techniques for Distributed Objects, Components, and Systems", Copenhagen, Denmark, J. A. PÉREZ, N. YOSHIDA (editors), Formal Techniques for Distributed Objects, Components, and Systems, Springer International Publishing, 2019, vol.

LNCS-11535, p. 224-241, Part 1: Full Papers [*DOI :* 10.1007/978-3-030-21759-4_13], https://hal.inria.fr/hal-02109769

[20] I. RAÏS, D. BALOUEK-THOMERT, A.-C. ORGERIE, L. LEFÈVRE, M. PARASHAR.*Leveraging energy-efficient non-lossy compression for data-intensive applications*, in "HPCS 2019 - 17th International Conference on High Performance Computing & Simulation", Dublin, Ireland, HPCS 2019 - 17th International Conference on High Performance Computing & Simulation, July 2019, p. 1-7, https://hal.archives-ouvertes.fr/hal-02179621

[21] G. TATO, M. BERTIER, E. RIVIÈRE, C. TEDESCHI.*Split and migrate: Resource-driven placement and discovery of microservices at the edge*, in "OPODIS 2019 : 23rd International Conference On Principles Of Distributed Systems", Neuchâtel, Switzerland, 2019, p. 1-16 [*DOI :* 10.4230/LIPICS.OPODIS.2019.29], https://hal.inria.fr/hal-02401933

[22] A. T. WONJIGA, S. PEISERT, L. RILLING, C. MORIN.*Blockchain as a Trusted Component in Cloud SLA Verification*, in "UCC 2019 - 12th IEEE/ACM International Conference on Utility and Cloud Computing", Auckland, New Zealand, ACM, December 2019, p. 93-100 [*DOI :* 10.1145/3368235.3368872], https://hal.inria.fr/hal-02354647

National Conferences with Proceeding

[23] C. ELBAZ, L. RILLING, C. MORIN.*Towards Automated Risk Analysis of "One-day" Vulnerabilities*, in "RESSI 2019 - Rendez-vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information", Erquy, France, August 2019, p. 1-3, https://hal.inria.fr/hal-02267192

Scientific Books (or Scientific Book chapters)

[24] A. OLEKSIAK, L. LEFÈVRE, P. ALONSO, G. DA COSTA, V. DE MAIO, N. FRASHERI, V. GARCIA, J. GUERRERO, S. LAFOND, A. LASTOVETSKY, R. R. MANUMACHU, B. MUITE, A.-C. ORGERIE, W. PIATEK, J.-M. PIERSON, R. PRODAN, P. STOLF, E. SHEME, S. VARRETTE.*Energy aware ultrascale systems*, in "Ultrascale Computing Systems", Institution of Engineering and Technology, January 2019, p. 127-188 [*DOI :* 10.1049/PBPC024E_CH], https://hal.inria.fr/hal-02163289

Research Reports

[25] C. ELBAZ, L. RILLING, C. MORIN.*Automated Keyword Extraction from "One-day" Vulnerabilities at Disclosure*, Inria Rennes - Bretagne Atlantique, November 2019, n° RR-9299, p. 1-22, https://hal.inria.fr/hal-02362062

[26] A. JLASSI, C. TEDESCHI.*Merge, Split, and Cluster: Dynamic Deployment of Stream Processing Applications*, Inria, December 2019, https://hal.inria.fr/hal-02418771

[27] A. T. WONJIGA, L. RILLING, C. MORIN.*Defining Security Monitoring SLAs in IaaS Clouds: the Example of a Network IDS*, Inria Rennes Bretagne Atlantique, March 2019, n° RR-9263, p. 1-37, https://hal.inria.fr/hal-02079860

[28] A. T. WONJIGA, L. RILLING, C. MORIN.*Security Monitoring SLA Verification in Clouds: the Case of Data Integrity*, Inria Rennes - Bretagne Atlantique, March 2019, n° RR-9267, p. 1-29, https://hal.inria.fr/hal-02084186

Other Publications

[29] M. BELKHIRIA, C. TEDESCHI.*Decentralized Scaling for Stream Processing Engines*, May 2019, working paper or preprint, https://hal.inria.fr/hal-02127609

[30] A. GUERMOUCHE, A.-C. ORGERIE.*Experimental analysis of vectorized instructions impact on energy and power consumption under thermal design power constraints*, June 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02167083

References in notes

[31] R.-A. CHERRUEAU, M. SIMONIN, A. VAN KEMPEN.*EnosStack: A LAMP-like stack for the experimenter*, in "INFOCOM WKSHPS 2018 - IEEE International Conference on Computer Communications", Honolulu, United States, IEEE, April 2018, p. 336-341 [*DOI :* 10.1109/INFCOMW.2018.8407024], https://hal.inria.fr/hal-01916484

[32] M. GUTHMULLER, G. CORONA, M. QUINSON.*System-level state equality detection for the formal dynamic verification of legacy distributed applications*, in "Journal of Logical and Algebraic Methods in Programming", April 2018, vol. 96, p. 1 - 11 [*DOI :* 10.1016/J.JLAMP.2017.12.004], https://hal.inria.fr/hal-01900120

[33] A. LEBRE, 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° RR-8348, https://hal.inria.fr/hal-00854204

[34] J. ROJAS BALDERRAMA, M. SIMONIN.*Scalability and Locality Awareness of Remote Procedure Calls: An Experimental Study in Edge Infrastructures*, in "CloudCom", Nicosia, Cyprus, December 2018, https://hal.inria.fr/hal-01891567

[35] G. TATO, M. BERTIER, E. RIVIÈRE, C. TEDESCHI.*ShareLatex on the Edge: Evaluation of the Hybrid Core/Edge Deployment of a Microservices-based Application*, in "The 3rd workshop on Middleware for Edge Clouds & Cloudlets", Rennes, France, December 2018, https://hal.inria.fr/hal-01942807

[36] G. TATO, M. BERTIER, C. TEDESCHI.*Koala: Towards Lazy and Locality-Aware Overlays for Decentralized Clouds*, in "ICFEC 2018 - 2nd IEEE International Conference on Fog and Edge Computing", Washington, United States, IEEE, May 2018, p. 1-10 [*DOI :* 10.1109/CFEC.2018.8358728], https://hal.inria.fr/hal-01938582

[37] 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, IEEE, April 2018, p. 1-7 [*DOI :* 10.1109/NOMS.2018.8406157], https://hal.inria.fr/hal-01663941

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

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
A2.2.4. - Parallel architectures
A2.2.5. - Run-time systems
A2.2.6. - GPGPU, FPGA...
A2.2.7. - Adaptive compilation
A2.2.8. - Code generation
A2.2.9. - Security by 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. Team, Visitors, External Collaborators

Research Scientists

Erven Rohou [Team leader, Inria, Senior Researcher, HDR]
Caroline Collange [Inria, Researcher]
Byron Hawkins [Inria, Starting Research Position]
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

Alexandre Kouyoumdjian [Inria, Engineer]
Stefanos Skalistis [Inria, Engineer, from Jul 2019]
Loïc Besnard [CNRS]

PhD Students

Nicolas Bellec [Inria, from Sep 2019]
Arthur Blanleuil [Univ de Rennes I]
Niloofar Charmchi [Inria]
Kleovoulos Kalaitzidis [Inria]
Kévin Le Bon [Inria]
Anis Peysieux [Inria, from Oct 2019]
Daniel Rodrigues Carvalho [Inria]
Bahram Yarahmadi [Inria]

Post-Doctoral Fellows

Pierre-Yves Peneau [Inria]
Stefanos Skalistis [Univ de Rennes I, until Jun 2019]

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 micro-architectures. 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 is experiencing the end of Moore's law [0], 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 of the PACAP project-team derives 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 micro-architecture.

2.1.4. Throughput-Oriented Computing

The goal is to maximize the performance-to-power ratio. We will leverage the execution model of throughput-oriented 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:

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.

[0] Moore's law states that the number of transistors in a circuit doubles (approximately) every two years.

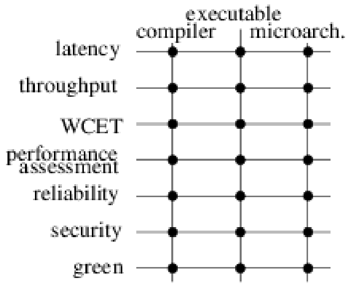


Figure 1. Perimeter of Research Objectives

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 workloads 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 micro-architecture, 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 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 micro-architecture and compilation techniques, we will be able to significantly improve efficiency, robustness and coverage of security mechanisms, as well as to partner with field experts to design innovative solutions.

2.1.8. Green Computing – Power Concerns

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 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 thus confronted with a moving target, challenging portability and jeopardizing performance.

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-Silicon, 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), quantum computing (impacts the entire software stack).

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 a 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, performance-oriented compilation and guaranteed response time for real-time. The research results may have impact 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 [40] 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 increased performance.

The end of Moore's law and Dennard's scaling [0] 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 micro-architecture (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 power efficiency (Section 3.2.6), and security (Section 3.2.7).

3.2.1. Static Compilation

Static compilation techniques continue to be relevant in addressing 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 new optimization opportunities and develops tailored compilation techniques for 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.

[0] According to Dennard scaling, as transistors get smaller the power density remains constant, and the consumed power remains proportional to the area.

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 [40]: 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 transparently [2]. 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 [46] 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 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.

3.2.3. Research directions in uniprocessor micro-architecture

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 micro-architecture 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 recent years they have been recognized as world leaders in branch prediction [50] [44] and in cache prefetching [6] and they have revived the forgotten concept of value prediction [9][8]. 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 cores;
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) [6]. The first research objective is to better understand how the implicit prefetching enabled by the large instruction window interacts with the L2 prefetcher and then to understand how explicit prefetching on the L1 also interacts with the L2 prefetcher.

The second research objective 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 benefits 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 [11]. It offers new possibilities 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 processors. For the two past decades, implementing ever wider issue superscalar processors has been challenging. The objective of our research on the execution core is to explore (and revisit) directions that 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 architectures [7]. Symmetric clustered organization allows to benefit from a simpler bypass network, but induce large complexity on the issue queue. One remarkable finding of our study [7] is that, when considering two large clusters (e.g. 8-wide), steering large groups of consecutive instructions (e.g. 64 μops) 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. [47]. 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-urgent. 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 micro-architecture with a cluster dedicated to the execution of urgent instructions and a second cluster executing the non-urgent instructions. The micro-architecture 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 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 [48].

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 [9][8] [50] [44]. First, very high accuracy can be enforced at reasonable cost in coverage and minimal complexity [9]. 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 [8]. 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 [43] 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 micro-architectures 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 compiler levels.

Architecture.
The architecture convergence between Single Instruction Multiple Threads (SIMT) GPUs and multicore processors that we have been pursuing [42] 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.

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 [0], or the Recore manycore hardware [0]) 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 analyze and/or control 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. 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 many-core 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.

[0]http://www.kalrayinc.com
[0]http://www.recoresystems.com/

3.2.7. 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 in the past, thanks to the HAVEGE [49] random number generator, and code obfuscating techniques (the obfuscating just-in-time compiler [41], or thread-based control flow mangling [45]). Still, security is not core competence of PACAP members.

Our strategy consists in partnering 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 expertise in compilation and architecture helps design more efficient and less expensive protection mechanisms.

Examples of collaborations so far include the following:

Compilation: We partnered with experts in security and codes to prototype a platform that demonstrates resilient software. They designed and proposed advanced masking techniques to hide sensitive data in application memory. 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.

Dynamic Binary Rewriting: Our expertise in dynamic binary rewriting combines well with the expertise of the CIDRE team in protecting application. Security has a high cost in terms of performance, and static insertion of counter measures cannot take into account the current threat level. In collaboration with CIDRE, we propose an adaptive insertion/removal of countermeasures in a running application based of dynamic assessment of the threat level.

WCET Analysis: Designing real-time systems requires computing an upper bound of the worst-case execution time. Knowledge of this timing information opens an opportunity to detect attacks on the control flow of programs. In collaboration with CIDRE, we are developing a technique to detect such attacks thanks to a hardware monitor that makes sure that statically computed time information is preserved (CAIRN is also involved in the definition of the hardware component).

4. Application Domains

4.1. Domains

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 impact 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. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

Benjamin Rouxel, Stefanos Skalistis, Steven Derrien and Isabelle Puaut received an Outstanding paper award for their paper entitled "Hiding Communication Delays in Contention-Free Execution for SPM-based Multi-Core Architectures" at the Euromicro conference on real time systems .
BEST PAPERS AWARDS :
[28]

B. ROUXEL, S. SKALISTIS, S. DERRIEN, I. PUAUT.*Hiding Communication Delays in Contention-Free Execution for SPM-Based Multi-Core Architectures*, in "ECRTS 2019 - 31st Euromicro Conference on Real-Time Systems", Stuttgart, Germany, July 2019, p. 1-24 [*DOI :* 10.4230/LIPICS.ECRTS.2019.25], https://hal.archives-ouvertes.fr/hal-02190271

6. New Software and Platforms

6.1. 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 time-varying thermal behavior of a processing chip.

ATMI (Analytical model of Temperature in MIcroprocessors) is an ad hoc temperature model 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/

6.2. HEPTANE

KEYWORDS: IPET - WCET - Performance - Real time - Static analysis - Worst Case Execution Time

SCIENTIFIC DESCRIPTION: WCET estimation

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.

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/

6.3. tiptop

KEYWORDS: Instructions - Cycles - Cache - CPU - Performance - HPC - Branch predictor

SCIENTIFIC DESCRIPTION: Tiptop is a new simple and flexible user-level tool that collects hardware counter data on Linux platforms (version 2.6.31+) and displays them in a way simple to the Linux "top" utility. The goal is to make the collection of performance and bottleneck data as simple as possible, including simple installation and usage. No privilege is required, any user can run tiptop.

Tiptop is written in C. It can take advantage of libncurses when available for pseudo-graphic display. 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.

Current version is 2.3.1, released October 2017. Tiptop has been integrated in major Linux distributions, such as Fedora, Debian, Ubuntu, CentOS.

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

6.4. 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/pacap/software/padrone

6.5. 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.

FUNCTIONAL DESCRIPTION: If-memo is a linker-based technique for enabling software memorizing of any dynamically linked pure function by function interception. Typically, this framework is useful to intercept the computationally expensive pure functions - the transcendental functions from the math library. 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/pacap/software/if-memo/

6.6. Simty

KEYWORDS: GPU - Softcore - FPGA - SIMT - Multi-threading - RISC-V

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: Caroline Collange
- Contact: Caroline Collange
- URL: https://gforge.inria.fr/projects/simty

6.7. Barra

KEYWORDS: GPU - GPGPU - Tesla ISA - Debug - Computer architecture - Performance - Profiling - Simulator - HPC - CUDA

SCIENTIFIC DESCRIPTION: Research on throughput-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: Version 0.5.10 introduces: Timing model, Tesla-like architecture model, Fermi-like architecture model, New per-PC control-flow divergence management, Support for Simultaneous branch and warp interweaving, Support for Affine vector cache.

- Participants: Alexandre Kouyoumdjian, David Defour, Fabrice Mouhartem and Caroline Collange
- Partners: ENS Lyon - UPVD
- Contact: Caroline Collange
- URL: http://barra.gforge.inria.fr/

6.8. Memoization

KEYWORDS: Optimization - Pure function - Memoization

FUNCTIONAL DESCRIPTION: Memoization is a technique used at runtime that consists in caching results of pure functions and retrieving them instead of computing them when the arguments repeat. It can be applied to C and C++ programs. To be memoized, the interface of a pure function (or a method) must verify the following properties: (1) the function/method has at most four arguments of same type T, (2) the function/method returns a data of type T, (3) T is either 'double', 'float', or 'int'.

The memoization operation of a function/method is controlled by several parameters: the size of the internal table (number of entries), the replacement policy to be used in case of index conflict (whether the value of the table must be replaced or not), an approximation threshold that allows to not distinguish very close values). It is also possible to initialize the table with the content of a file, and to save the content of the table to a file at the end of the execution (the data may be used as input for a future execution).

- Participants: Loïc Besnard, Imane Lasri and Erven Rohou
- Contact: Loïc Besnard

6.9. FiPlib

KEYWORDS: Compilation - Approximate computing - Fixed-point representation

FUNCTIONAL DESCRIPTION: FiPlib is a C++ library that provides type definition and conversion operations for computations in fixed-point representation. Basic arithmetic as well as logical operations are transparently supported thanks to operator overloading. FiPlib also provides optimized implementations of the transcendental math functions of libm. For convenient integration, FiPlib is released as C++ header files only. Optionally, FiPlib can detect overflows and compute errors compared to floating point representation.

- Participants: Pierre Le Meur, Imane Lasri and Erven Rohou
- Contact: Erven Rohou

6.10. sigmask

KEYWORDS: Compilation - Side-channel - Masking - Security - Embedded systems

SCIENTIFIC DESCRIPTION: Sigmask is a compiler plugin based on the LLVM infrastructure that automatically protects secret information in programs, such as encryption keys, against side-channel attacks. The programmer annotates their source code to highlight variables containing sensitive data. The compiler automatically analyzes the program and computes all memory locations potentially derived from the secret. It then applies a masking scheme to avoid information leakage. Sigmask provides several schemes: OSDM (Orthogonal Direct Sum Masking), IP (Inner Product) Masking, and simple random bit masking. The programmer may also provide their own masking scheme through a well-defined API.

FUNCTIONAL DESCRIPTION: Sigmask is a compiler plugin based on the LLVM infrastructure that automatically protects secret information in programs, such as encryption keys, against side-channel attacks. The programmer annotates their source code to highlight variables containing sensitive data. The compiler automatically analyzes the program and computes all memory locations potentially derived from the secret. It then applies a masking scheme to avoid information leakage. Sigmask provides several schemes: ODSM (Orthogonal Direct Sum Masking), IP (Inner Product) Masking, and simple random bit masking. The programmer may also provide their own masking scheme through a well-defined API.

- Participants: Nicolas Kiss, Damien Hardy and Erven Rohou
- Contact: Erven Rohou

7. New Results

7.1. Compilation and Optimization

Participants: Loïc Besnard, Caroline Collange, Byron Hawkins, Erven Rohou, Bahram Yarahmadi.

7.1.1. Optimization in the Presence of NVRAM

Participants: Erven Rohou, Bahram Yarahmadi.

A large and increasing number of Internet-of-Things devices are not equipped with batteries and harvest energy from their environment. Many of them cannot be physically accessed once they are deployed (embedded in civil engineering structures, sent in the atmosphere or deep in the oceans). When they run out of energy, they stop executing and wait until the energy level reaches a threshold. Programming such devices is challenging in terms of ensuring memory consistency and guaranteeing forward progress.

7.1.1.1. Checkpoint Placement based Worst-Case Energy Consumption

Previous work has proposed to insert checkpoints in the program so that execution can resume from well-defined locations. We propose to define these checkpoint locations based on worst-case energy consumption of code sections, with limited additional effort for programmers. As our method is based upon worst-case energy consumption, we can guarantee memory consistency and forward progress.

This work has been presented at the Compas 2019 conference.

7.1.1.2. Dynamic Adaptive Checkpoint Placement

Previous work has proposed to back-up the volatile states which are necessary for resuming the program execution after power failures. They either do it at compile time by placing checkpoints into the control flow of the program or at runtime by leveraging voltage monitoring facilities and interrupts, so that execution can resume from well-defined locations after power failures. We propose for the first time a dynamic checkpoint placement strategy which delays checkpoint placement and specialization to the runtime and takes decisions based on the past power failures and execution paths that are taken. We evaluate our work on a TI MSP430 device, with different types of benchmarks as well as different uninterrupted intervals, and we measure the execution time. We show that our work can outperform compiler-based state-of-the-art with memory footprint kept under the control.

This research is done within the context of the project IPL ZEP.

7.1.2. Dynamic Binary Optimization

Participant: Erven Rohou.

7.1.2.1. Guided just-in-time specialization

JavaScript's portability across a vast ecosystem of browsers makes it today a core building block of the web. Yet, building efficient systems in JavaScript is still challenging. Because this language is so dynamic, JavaScript programs provide little information that just-in-time compilers can use to carry out safe optimizations. Motivated by this observation, we propose to guide the JIT compiler in the task of code specialization. To this end, we have augmented [17] the language with an annotation that indicates which function call sites are likely to benefit from specialization. To support the automatic annotation of programs, we have introduced a novel static analysis that identifies profitable specialization points. We have implemented our ideas in JavaScriptCore, the built-in JavaScript engine for WebKit. The addition of guided specialization to this engine required us to change it in several non-trivial ways. Such changes let us observe speedups of up to 1.7× on programs present in synthetic benchmarks.

7.1.2.2. Run-time parallelization and de-parallelization

Runtime compilation has opportunities to parallelize code which are generally not available using static parallelization approaches. However, the parallelized code can possibly slowdown the performance due to unforeseen parallel overheads such as synchronization and speculation support pertaining to the chosen parallelization strategy and the underlying parallel platform. Moreover, with the wide usage of heterogeneous architectures, such choice options become more pronounced. We consider [22] an adaptive form of the parallelization operation, for the first time. We propose a method for performing on-stack de-parallelization for a parallelized binary loop at runtime, thereby allowing for rapid loop replacement with a more optimized one. We consider a loop parallelization strategy and propose a corresponding de-parallelization method. The method relies on stopping the execution at safe points, gathering threads' states, producing a corresponding serial code, and continuing execution serially. The decision to de-parallelize or not is taken based on the anticipated speedup. To assess the extent of our approach, we have conducted an initial study on a small set of programs with various parallelization overheads. Results show up to 4× performance improvement for a synchronization intense program on a 4-core Intel processor.

With the multicore trend, the need for automatic parallelization is more pronounced, especially for legacy and proprietary code where no source code is available and/or the code is already running and restarting is not an option. We engineer [21] a mechanism for transforming at runtime a frequent for-loop with no data dependencies in a binary program into a parallel loop, using on-stack replacement. With our mechanism, there is no need for source code, debugging information or restarting the program. Also, the mechanism needs no static instrumentation or information. The mechanism is implemented using the Padrone binary modification system and `pthreads`, where the remaining iterations of the loop are executed in parallel. The mechanism keeps the running program state by extracting the targeted loop into a separate function and copying the current stack frame into the corresponding frames of the created threads. Initial study is conducted on a set of kernels from the Polybench workload. Experimental results show from 2× to 3.5× speedup from sequential to parallelized code on four cores, which is similar to source code level parallelization.

This research was partially done within the context of the project PHC IMHOTEP.

7.1.3. Automatic and Parametrizable Memoization

Participants: Loïc Besnard, Erven Rohou.

Improving execution time and energy efficiency is needed for many applications and usually requires sophisticated code transformations and compiler optimizations. One of the optimization techniques is memoization, which saves the results of computations so that future computations with the same inputs can be avoided. We propose [16] a framework that automatically applies memoization techniques to C/C++ applications. The framework is based on automatic code transformations using a source-to-source compiler and on a memoization library. With the framework users can select functions to memoize as long as they obey to certain restrictions imposed by our current memoization library. We show the use of the framework and associated memoization technique and the impact on reducing the execution time and energy consumption of four representative benchmarks. The support library is available at https://gforge.inria.fr/projects/memoization (registered with APP under number IDDN.FR.001.250029.000.S.P.2018.000.10800).

7.1.4. Autotuning

Participants: Loïc Besnard, Erven Rohou.

The ANTAREX FET HPC project relies on a Domain Specific Language (DSL) based on Aspect Oriented Programming (AOP) concepts to allow applications to enforce extra functional properties such as energy-efficiency and performance and to optimize Quality of Service (QoS) in an adaptive way. The DSL approach allows the definition of energy-efficiency, performance, and adaptivity strategies as well as their enforcement at runtime through application autotuning and resource and power management. We present [20] an overview of the key outcome of the project, the ANTAREX DSL, and some of its capabilities through a number of examples, including how the DSL is applied in the context of the project use cases. We demonstrated [30] tools and techniques in two domains: computational drug discovery, and online vehicle navigation.

7.1.5. Loop splitting

The loop splitting technique takes advantage of long running loops to explore the impact of several optimization sequences at once, thus reducing the number of necessary runs. We rely on a variant of loop peeling which splits a loop into into several loops, with the same body, but a subset of the iteration space. New loops execute consecutive chunks of the original loop. We then apply different optimization sequences on each loop independently. Timers around each chunk observe the performance of each fragment. This technique may be generalized to combine compiler options and different implementations of a function called in a loop. It is useful when, for example, the profiling of the application shows that a function is critical in term of time of execution. In this case, the user must try to find the best implementation of their algorithm.

This research was partially done within the context of the ANTAREX FET HPC collaborative project, collaboration is currently ongoing with University of Porto, Portugal.

7.1.6. Hardware/Software JIT Compiler

Participant: Erven Rohou.

Single-ISA heterogeneous systems (such as ARM big.LITTLE) are an attractive solution for embedded platforms as they expose performance/energy trade-offs directly to the operating system. Recent works have demonstrated the ability to increase their efficiency by using VLIW cores, supported through Dynamic Binary Translation (DBT) to maintain the illusion of a single-ISA system. However, VLIW cores cannot rival with Out-of-Order (OoO) cores when it comes to performance, mainly because they do not use speculative execution. We study [27] how it is possible to use memory dependency speculation during the DBT process. Our approach enables fine-grained speculation optimizations thanks to a combination of hardware and software. Our results show that our approach leads to a geo-mean speed-up of 10 % at the price of a 7 % area overhead.

Our previous work on Hybrid-DBT was also presented at the RISC-V workshop in Zürich, Switzerland [38].

This work is a collaboration with the CAIRN team.

7.1.7. Scalable program tracing

Participants: Byron Hawkins, Erven Rohou.

The initial goal of scalable tracing is to record long executions at under 5× overhead (ideally 2×), but it is equally important for analysis of the compressed trace to be efficient. This requires careful organization of the recorded data structures so that essential factors can be accessed without decompressing the trace or comprehensively iterating its paths. Precise context sensitivity is especially important for both optimization and security applications of trace-based program analysis, but scalability becomes challenging for frequently invoked functions that have a high degree of internal complexity. To avoid state space explosion in the context graph, such a function can be represented as a singleton while its complexity is preserved orthogonally. The current efforts focus mainly on developing an integration strategy to simplify program analysis over these two orthogonal dimensions of the trace.

7.1.8. Compiler optimization for quantum architectures

Participant: Caroline Collange.

In 2016, the first quantum processors have been made available to the general public. The possibility of programming an actual quantum device has elicited much enthusiasm [34]. Yet, such possibility also brought challenges. One challenge is the so called Qubit Allocation problem: the mapping of a virtual quantum circuit into an actual quantum architecture. There exist solutions to this problem; however, in our opinion, they fail to capitalize on decades of improvements on graph theory.

In collaboration with the Federal University of Minas Gerais, Brazil, we show how to model qubit allocation as the combination of Subgraph Isomorphism and Token Swapping [31]. This idea has been made possible by the publication of an approximative solution to the latter problem in 2016. We have compared our algorithm against five other qubit allocators, all independently designed in the last two years, including the winner of the IBM Challenge. When evaluated in “Tokyo”, a quantum architecture with 20 qubits, our technique outperforms these state-of-the-art approaches in terms of the quality of the solutions that it finds and the amount of memory that it uses, while showing practical runtime.

7.2. Processor Architecture

Participants: Arthur Blanleuil, Niloofar Charmchi, Caroline Collange, Kleovoulos Kalaitzidis, Pierre Michaud, Anis Peysieux, Daniel Rodrigues Carvalho, André Seznec.

7.2.1. Value prediction

Participants: Kleovoulos Kalaitzidis, André Seznec.

Modern context-based value predictors tightly associate recurring values with instructions and contexts by building confidence upon them [9]. However, when execution monotony exists in the form of intervals, the potential prediction coverage is limited, since prediction confidence is reset at the beginning of each new interval. In [25], we address this challenge by introducing the notion of Equality Prediction (EP), which represents the binary facet of value prediction. Following a two fold decision scheme (similar to branch prediction), EP makes use of control-flow history to determine equality between the last committed result read at fetch time, and the result of the fetched occurrence. When equality is predicted with high confidence, the read value is used. Our experiments show that this technique obtains the same level of performance as previously proposed state-of-the-art context-based predictors. However, by virtue of better exploiting patterns of interval equality, our design complements the established way that value prediction is performed, and when combined with contemporary prediction models, improves the delivered speedup by 19 % on average.

7.2.2. Compressed caches

Participants: Daniel Rodrigues Carvalho, Niloofar Charmchi, Caroline Collange, André Seznec.

The speed gap between CPU and memory is impairing performance. Cache compression and hardware prefetching are two techniques that could confront this bottleneck by decreasing last level cache misses. However, compression and prefetching have positive interactions, as prefetching benefits from higher cache capacity and compression increases the effective cache size. We propose Compressed cache Layout Aware Prefetching (CLAP) to leverage the recently proposed sector-based compressed cache layouts such as SCC or YACC to create a synergy between compressed cache and prefetching. The idea of this approach is to prefetch contiguous blocks that can be compressed and co-allocated together with the requested block on a miss access [33]. Prefetched blocks that share storage with existing blocks do not need to evict a valid existing entry; therefore, CLAP avoids cache pollution. In order to decide the co-allocatable blocks to prefetch, we propose a compression predictor. Based on our experimental evaluations, CLAP reduces the number of cache misses by 12 % and improves performance by 4 % on average, comparing to a compressed cache [23].

7.2.3. Deep microarchitecture

Participants: Anis Peysieux, André Seznec.

The design of an efficient out-of-order execution core is particularly challenging. When the issue-width increases, the cost of the extra logic required by out-of-core execution increases dramatically. The silicon area occupied by this OoO core tends to grow quasi-quadratically with the issue-width (e.g. issue logic, register file and result bypass). At the same time, the power requirement and the energy consumption of the out-of–order core grow super-linearly with issue width. On wide-issue out-of-order execution cores, issue logic response time, register file access time, as well as result bypass delays represent potential critical paths that might impair cycle time or might necessitate further deepening of the execution pipeline. The objective of the PhD thesis of Anis Peysieux will be to reduce the number of instructions that enter the OoO core, and therefore to master the hardware complexity while still achieving the performance promises of a very wide issue processor.

7.2.4. Dynamic thermal management

Participant: Pierre Michaud.

As power dissipation and circuit temperature constrain their performance, modern processors feature turbo control mechanisms to adjust the voltage and clock frequency dynamically so that circuit temperature stays below a certain limit. In particular, turbo control exploits the fact that, after a long period of low processor activity, the thermal capacity of the chip, its package and the heatsink can absorb heat at a relatively fast rate during a certain time, before the temperature limit constrains that rate. Hence power dissipation can be temporarily boosted above the average sustainable value. The turbo control must monitor circuit temperature continuously to maximize the clock frequency. Temperature can be monitored by reading the integrated thermal sensors. However, making the clock frequency depend on thermal sensor readings implies that processor performance depends on ambient temperature. Yet this form of performance non-determinism is a problem for certain processor makers. A possible solution is to determine the clock frequency not from the true temperature but from a thermal model based on the nominal ambient temperature. Such model should be as accurate as possible in order to prevent sensor-based protection from triggering but sporadically, without hurting performance by overestimating temperature too much. The model should also be simple enough to provide calculated temperature in real time. We propose a thermal model possessing these qualities, and a new turbo control algorithm based on that model [37].

7.2.5. Thread convergence prediction for general-purpose SIMT architectures

Participants: Arthur Blanleuil, Caroline Collange.

GPUs group threads of SPMD programs in warps and synchronize them to execute the same instruction at the same time. This execution model, referred to as Single-Instruction, Multiple-Thread (SIMT), enables the use of energy-efficient SIMD execution units by factoring out control logic such as instruction fetch and decode pipeline stages for a whole warp. SIMT execution is the key enabler for the energy efficiency of GPUs. We seek to generalize the SIMT execution model to general-purpose superscalar cores.

As threads within a warp may follow different directions through conditional branches in the program, the warp must follow each taken path in turn, while disabling individual threads that do not participate. Following divergence, current GPU architectures attempt to restore convergence at the earliest program point following static annotations in the binary. However, this policy has been shown to be suboptimal in many cases, in which later convergence improves performance. In fact, optimal convergence points depend on dynamic program behavior, so static decisions are unable to capture them.

The goal of the thesis of Arthur Blanleuil is to design predictors that enable the microarchitecture to infer dynamic code behavior and place convergence points appropriately. Convergence predictors have analogies with branch predictors and control independence predictors studied in superscalar processor architecture, but they present one additional challenge: the thread runaway problem. Although a branch misprediction will be identified and repaired locally, a wrong thread scheduling decision may go unnoticed and delay convergence by thousands of instructions. To address the thread runaway problem, we plan to explore promise-based speculation and recovery strategies. When no information is available, we follow the traditional conservative earliest-convergence scheduling policy. Once the predictor has enough information to make a more aggressive prediction, it generates assumptions about the prediction. The microarchitecture then keeps checking dynamically whether the assumptions actually hold true in the near future. If assumptions turn out to be wrong, the prediction will be reconsidered by changing back priorities to conservative. Such promise-based speculation policies can address the thread runaway problem by fixing a bound on the worst-case performance degradation of an aggressive scheduling policy against the conservative baseline.

Accurate thread convergence policies will enable dynamic vectorization to adapt to application characteristics dynamically. They will both improve performance and simplify programming of many-core architectures by alleviating the need for advanced code tuning by expert programmers.

7.2.6. Exploring the design space of GPU architectures

Participants: Alexandre Kouyoumdjian, Caroline Collange.

We study tradeoffs in the internal organization of GPUs in the context of general-purpose parallel processing [35]. In particular, we analyze the performance impact of having a few wide streaming multiprocessors compared to many narrow ones. Although we find narrow configurations usually give higher performance for an equal number of execution units, they require more hardware resources and energy. On the other hand, our evaluation show that the optimal streaming multiprocessor width varies across applications. This study motivates adaptive GPU architectures that would support configurable internal organization.

7.3. WCET estimation and optimization

Participants: Loïc Besnard, Damien Hardy, Isabelle Puaut, Stefanos Skalistis.

7.3.1. WCET estimation for many core processors

Participants: Damien Hardy, Isabelle Puaut, Stefanos Skalistis.

7.3.1.1. 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 is 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 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 co-locating 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. An implementation on the Kalray MPPA machine allowed to identify implementation-related overheads. All results are described in [18].

7.3.1.2. Shared resource contentions and WCET estimation

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 [28], we propose a scheduling technique that jointly selects Scratchpad Memory (SPM) contents off-line, in such a way that the cost of SPM loading/unloading is hidden. Communications are fragmented to augment hiding possibilities. Experimental results show the effectiveness of the proposed technique on streaming applications and synthetic task-graphs. The overlapping of communications with computations allows the length of generated schedules to be reduced by 4 % on average on streaming applications, with a maximum of 16 %, and by 8 % on average for synthetic task graphs. We further show on a case study that generated schedules can be implemented with low overhead on a predictable multi-core architecture (Kalray MPPA).

7.3.1.3. Interference-sensitive run-time adaptation of time-triggered schedules

In time-critical systems, run-time adaptation is required to improve the performance of time-triggered execution, derived based on Worst-Case Execution Time (WCET) of tasks. By improving performance, the systems can provide higher Quality-of-Service, in safety-critical systems, or execute other best-effort applications, in mixed-critical systems. To achieve this goal, we propose in [32] a parallel interference-sensitive run-time adaptation mechanism that enables a fine-grained synchronisation among cores. Since the run-time adaptation of offline solutions can potentially violate the timing guarantees, we present the Response-Time Analysis (RTA) of the proposed mechanism showing that the system execution is free of timing-anomalies. The RTA takes into account the timing behavior of the proposed mechanism and its associated WCET. To support our contribution, we evaluate the behavior and the scalability of the proposed approach for different application types and execution configurations on the 8-core Texas Instruments TMS320C6678 platform. The obtained results show significant performance improvement compared to state-of-the-art centralized approaches.

7.3.1.4. WCET-Aware Parallelization of Model-Based Applications for Multi-Cores

Parallel architectures are nowadays not only confined to the domain of high performance computing, they are also increasingly used in embedded time-critical systems.

The 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. The Argo toolchain operates on Scilab and XCoS inputs, and targets ScratchPad Memory (SPM)-based multi-cores. Data-layout and loop transformations play a key role in this flow as they improve SPM efficiency and reduce the number of accesses to shared main memory.

In [19] we present the overall results of the project, a compiler tool-flow for automated parallelization of model-based real-time software, which addresses the shortcomings of multi-core architectures in real-time systems. The flow is demonstrated using a model-based Terrain Awareness and Warning Systems (TAWS) and an edge detection algorithm from the image-processing domain. Model-based applications are first transformed into real-time C code and from there into a well-predictable parallel C program. Tight bounds for the Worst-Case Execution Time (WCET) of the parallelized program can be determined using an integrated multi-core WCET analysis. Thanks to the use of an architecture description language, the general approach is applicable to a wider range of target platforms. An experimental evaluation for a research architecture with network-on-chip (NoC) interconnect shows that the parallel WCET of the TAWS application can be improved by factor 1.77 using the presented compiler tools.

7.3.2. WCET estimation and optimizing compilers

Participants: Isabelle Puaut, Stefanos Skalistis.

Static Worst-Case Execution Time (WCET) estimation techniques operate upon the binary code of a program in order to provide the necessary input for schedulability analysis techniques. Compilers used to generate this binary code include tens of optimizations, that can radically change the flow information of the program. Such information is hard to be maintained across optimization passes and may render automatic extraction of important flow information, such as loop bounds, impossible. Thus, compiler optimizations, especially the sophisticated optimizations of mainstream compilers, are typically avoided. We explore [24] for the first time iterative-compilation techniques that reconcile compiler optimizations and static WCET estimation. We propose a novel learning technique that selects sequences of optimizations that minimize the WCET estimate of a given program. We experimentally evaluate the proposed technique using an industrial WCET estimation tool (AbsInt aiT) over a set of 46 benchmarks from four different benchmarks suites, including reference WCET benchmark applications, image processing kernels and telecommunication applications. Experimental results show that WCET estimates are reduced on average by 20.3 % using the proposed technique, as compared to the best compiler optimization level applicable.

7.3.3. WCET estimation and processor micro-architecture

Participant: Isabelle Puaut.

Cache memories in modern embedded processors are known to improve average memory access performance. Unfortunately, they are also known to represent a major source of unpredictability for hard real-time workload. One of the main limitations of typical caches is that content selection and replacement is entirely performed in hardware. As such, it is hard to control the cache behavior in software to favor caching of blocks that are known to have an impact on an application's worst-case execution time (WCET). In [26], we consider a cache replacement policy, namely DM-LRU, that allows system designers to prioritize caching of memory blocks that are known to have an important impact on an application's WCET. Considering a single-core, single-level cache hierarchy, we describe an abstract interpretation-based timing analysis for DM-LRU. We implement the proposed analysis in a self-contained toolkit and study its qualitative properties on a set of representative benchmarks. Apart from being useful to compute the WCET when DM-LRU or similar policies are used, the proposed analysis can allow designers to perform WCET impact-aware selection of content to be retained in cache.

Long pipelines need good branch predictors to keep the pipeline running. Current branch predictors are optimized for the average case, which might not be a good fit for real-time systems and worst-case execution time analysis. We present [29] a time-predictable branch predictor co-designed with the associated worst-case execution time analysis. Thee branch predictor uses a fully-associative cache to track branch outcomes and destination addresses. The fully-associative cache avoids any false sharing of entries between branches. Therefore, we can analyze program scopes that contain a number of branches lower than or equal to the number of branches in the prediction table. Experimental results show that the worst-case execution time bounds of programs using the proposed predictor are lower than using static branch predictors at a moderate hardware cost.

7.4. Security

Participants: Nicolas Bellec, Damien Hardy, Kévin Le Bon, Isabelle Puaut, Erven Rohou.

7.4.1. Attack detection co-processor for real-time systems

Participants: Nicolas Bellec, Isabelle Puaut.

Real-time embedded systems (RTES) are required to interact more and more with their environment, thereby increasing their attack surface. Recent security breaches on car brakes and other critical components, have already proven the feasibility of attacks on RTES. Such attacks may change the control-flow of the programs, which may lead to violations of the timing constraints of the system. In this ongoing work, we design a technique to detect attacks in RTES based on timing information. Our technique is based on a monitor, implemented in hardware to preserve the predictability of instrumented programs. The monitor uses timing information (Worst-Case Execution Time – WCET – of code regions) to detect attacks. An algorithm for the region selection, optimal when the monitoring memory is not limited is presented and provides guarantees on attack detection latency. An implementation of the hardware monitor and its simulation demonstrates the practicality of our approach. An experimental study evaluates the maximum attack detection latency for different monitor memory budgets.

This work is done in collaboration with the CIDRE and CAIRN teams.

7.4.2. Multi-nop fault injection attack

Participants: Damien Hardy, Erven Rohou.

The CIDRE team has developed a platform named Traitor that allows to perform multiple fault injection attack by replacing instructions by nops during the execution of a program. In this context, we are defining a program model where each instruction can be replaced by a nop at runtime. On this model we plan to apply compilation techniques on the binary to automatically determine where nops have to be inserted at runtime to perform sophisticated attacks such as dump of memory, modification of the memory, memory protection deactivation, execution of code in RAM.

This work is done in collaboration with the CIDRE team.

7.4.3. Compiler-based automation of side-channel countermeasures

Participants: Damien Hardy, Erven Rohou.

Masking is a popular protection against side-channel analysis exploiting the power consumption or electromagnetic radiations. Besides the many schemes based on simple Boolean encoding, some alternative schemes such as Orthogonal Direct Sum Masking (ODSM) or Inner Product Masking (IP) aim to provide more security, reduce the entropy or combine masking with fault detection. The practical implementation of those schemes is done manually at assembly or source-code level, some of them even stay purely theoretical. We proposed a compiler extension to automatically apply different masking schemes for block cipher algorithms. We introduced a generic approach to describe the schemes and we inserted three of them at compile-time on an AES implementation. Currently, a practical side-channel analysis is performed in collaboration with TAMIS to assess the correctness and the performance of the code inserted.

This work is done in collaboration with the TAMIS team.

7.4.4. Platform for adaptive dynamic protection of programs

Participants: Kévin Le Bon, Erven Rohou.

Memory corruption attacks are a serious threat for system integrity. Many techniques have been developed in order to protect systems from these attacks. However, the deployment of heavy protections often degrades the performance of programs. We propose [36] a dynamic approach that adapts the protection level of the target process during its execution depending on the observed behavior.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Grants with Industry

8.1.1. Intel research grant INTEL2016-11174

Participants: Niloofar Charmchi, Kleovoulos Kalaitzidis, Anis Peysieux, André Seznec.

Intel is supporting the research of the PACAP project-team on “Design tradeoffs for extreme cores”.

9. Partnerships and Cooperations

9.1. Regional Initiatives

The Brittany Region is partially funding the PhD fellowship for Niloofar Charmchi on the topic “Hardware prefetching and related issues” and Nicolas Bellec on the topic “Security in real-time embedded systems”.

9.2. National Initiatives

9.2.1. Zero Power Computing Systems (ZEP): Inria Project Lab (2017–2020)

Participants: Erven Rohou, Bahram Yarahmadi.

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 computing 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 [39]. The main application target is Internet of Things (IoT).

9.2.2. NOPE

Participants: Piéric Giraud, Erven Rohou, Bahram Yarahmadi.

NOPE is a one-year exploratory action funded by the Labex Cominlabs. This project aimed at being a first step, and served to elaborate more ambitious future works. Through this project, the consortium was able to grow its knowledge on a topical research theme and lay the foundations of an innovative hardware-software approach. The short term goals were:

- building and sharing across the consortium a strong expertise in state-of-the art results and tools on transient computing, and identifying challenges that should be focused on;
- initiating collaborations between the participants in order to identify opportunities at the hardware-software interface;
- building the foundations of a shared experimental platform for transient computing.

An intern, Piéric Giraud, was hired thanks to NOPE. He ported our WCET infrastructure Heptane to the MSP430 instruction set.

The NOPE project gathers teams PACAP, IETR Syscom and LS2N STR.

9.2.3. Hybrid SIMD architectures (2018–2019)

Participants: Caroline Collange, Alexandre Kouyoumdjian, Erven Rohou.

The project objective is to define new parallel computer architectures that offer high parallel performance on high-regularity workloads while keeping the flexibility to run more irregular parallel workloads. inspired by both GPU and SIMD or vector architectures.

This project is funded by the French Ministry of Armed Forces (*Ministère des Armées*).

9.2.4. DGA/PEC ARMOUR (2018–2021)

Participants: Kévin Le Bon, Erven Rohou.

ARMOUR (dynAmic binaRy optiMizatiOn cyber-secURity) aims at improving the security of computing systems at the software level. Our contribution will be twofold: (1) identify vulnerabilities in existing software, and (2) develop adaptive countermeasure mechanisms against attacks. We will rely on dynamic binary rewriting (DBR) which consists in observing a program and modifying its binary representation in memory while it runs. DBR does not require the source code of the programs it manipulates, making it convenient for commercial and legacy applications. We will study the feasibility of an adaptive security agent that monitors target applications and deploys (or removes) countermeasures based on dynamic conditions. Lightweight monitoring is appropriate when the threat condition is low, heavy countermeasures will be dynamically woven into the code when an attack is detected. Vulnerability analysis will be based on advanced fuzzing. DBR makes it possible to monitor and modify deeply embedded variables, inaccessible to traditional monitoring systems, and also to detect unexpected/suspicious values taken by variables and act before the application crashes.

ARMOUR is funded by DGA (*Direction Générale de l'Armement*) and PEC (*Pôle d'Excellence Cyber*).

9.2.5. ANR DYVE (31/03/2020 – 30/09/2023)

Participants: Arthur Blanleuil, Caroline Collange, Pierre-Yves Peneau.

Most of today's computer systems have CPU cores and GPU cores on the same chip. Though both are general-purpose, CPUs and GPUs still have fundamentally different software stacks and programming models, starting from the instruction set architecture. Indeed, GPUs rely on static vectorization of parallel applications, which demands vector instruction sets instead of CPU scalar instruction sets. In the DYVE project, we advocate a disruptive change in both CPU and GPU architecture by introducing Dynamic Vectorization at the hardware level.

Dynamic Vectorization will combine the efficiency of GPUs with the programmability and compatibility of CPUs by bringing them together into heterogeneous general-purpose multicores. It will enable processor architectures of the next decades to provide (1) high performance on sequential program sections thanks to latency-optimized cores, (2) energy-efficiency on parallel sections thanks to throughput-optimized cores, (3) programmability, binary compatibility and portability.

DYVE is funded by the ANR through the JCJC funding instrument.

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

9.3.1.1. ARGO

Participants: Damien Hardy, Isabelle Puaut, Stefanos Skalistis.

Title: Argo: WCET-Aware Parallelization of Model-Based Applications for Heterogeneous Parallel Systems

Program: H2020

Type: RIA

Duration: Jan 2016 – Mar 2019

Coordinator: Karlsruhe Institut für Technologie (Germany)

Université de Rennes 1 contact: Steven Derrien

Partners:

Karlsruher Institut für Technologie (Germany)

SCILAB enterprises SAS (France)

Université de Rennes 1 (France)

Technologiko Ekpaideftiko Idryma (TEI) Dytikis Elladas (Greece)

Absint GmbH (Germany)

Deutsches Zentrum für Luft- und Raumfahrt EV (Germany)

Fraunhofer (Germany)

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 many-core 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.

9.3.1.2. HiPEAC4 NoE

Participants: Pierre Michaud, Erven Rohou, André Seznec, Isabelle Puaut.

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.

9.3.1.3. Eurolab-4-HPC

Participant: Erven Rohou.

Title: EuroLab-4-HPC: Foundations of a European Research Center of Excellence in High Performance Computing Systems

Program: H2020

Duration: September 2018 – September 2020

Coordinator: Chalmers Tekniska Hoegskola AB (Sweden)

Partners:

Barcelona Supercomputing Center - Centro Nacional de Supercomputacion (Spain)

Chalmers Tekniska Hoegskola (Sweden)

Foundation for Research and Technology Hellas (Greece)

Universität Stuttgart (Germany)

The University of Manchester (United Kingdom)

Inria (France)

Universität Augsburg (Germany)

ETH Zürich (Switzerland)

École Polytechnique Federale de Lausanne (Switzerland)

Technion - Israel Institute of Technology (Israel)

The University of Edinburgh (United Kingdom)

Rheinisch-Westfaelische Technische Hochschule Aachen (Germany)

Universiteit Gent (Belgium)

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 long-term 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.

9.4. International Initiatives

9.4.1. ANR CHIST-ERA SECODE 2016–2019

Participants: Damien Hardy, Erven Rohou.

Title: SECODE – Secure Codes to Thwart Cyber-Physical Attacks

CHIST-ERA - RTCPS

Duration: January 2016 – December 2019 (one year extension)

Coordinator: Télécom Paris Tech (France)

Partners:

Télécom Paris Tech (France)

Inria (France)

Université Paris 8 (France)

Sabancı Üniversitesi (Turkey)

Université Catholique de Louvain (Belgium)

Inria contact: 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.

9.4.2. Informal International Partners

Caroline Collange has collaborated with Marcos Yukio Siraichi, Vinicius Fernandes dos Santos and Fernando Magno Quintão Pereira from UFMG, Brazil [31].

Isabelle Puaut has collaborated with Renato Mancuso (University of Boston, USA) and Heechul Yun (University of Kansas, USA) on predictable memory hierarchies [26]. She has collaborated with Martin Schoeberl (Technical University of Denmark) on predictable branch predictors [29].

Erven Rohou has been collaborating with Prof. Ahmed El-Mahdy (Egypt-Japan University of Science and Technology, Alexandria, Egypt) and his group [21], [22].

Erven Rohou and Loïc Besnard have been collaborating with Prof. João Cardoso (University of Porto, Porto, Portugal) and his group [16].

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Selection

10.1.1.1. Member of the Conference Program Committees

- Caroline Collange is a member of the program committee of the DATE 2020 conference
- Pierre Michaud was a member of the program committee of the International Conference on Computer Design (ICCD 2019)
- Pierre Michaud was a member of the program committee of the Third Data Prefetching Championship (DPC3)
- Isabelle Puaut was a member of the program committee of the Euromicro Conference on Real Time Systems (ECRTS) 2019 and 2020, IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS) 2020, Real Time Systems Symposium (RTSS) in 2019.
- Isabelle Puaut was a member of the program committee of the “Real-Time and (Networked) Embedded Systems” track of IEEE ETFA 2019.
- Isabelle Puaut was a member of the program committee of the International Conference on Real-Time Networks and Systems (RTNS) 2019 and 2020.
- Isabelle Puaut was a member of the program committee of the 18th Workshop on Worst-Case Execution Time Analysis (WCET 2019).
- Erven Rohou and Caroline Collange were members of the program committee of the French *Conférence francophone en informatique autour des thématiques du parallélisme, de l’architecture et des systèmes* (Compas 2019).
- Erven Rohou was a member of the program committee of the Special Session on Compiler Architecture, Design and Optimization (CADO) of HPCS 2019.
- André Seznec was a member of program committee of ACM/IEEE Micro 2019 conference
- André Seznec is a member of program committee of ACM/IEEE ISCA 2020 conference
- André Seznec is a member of program committee of the IPDPS 2020 conference

10.1.1.2. Reviewer

Members of PACAP routinely review submissions to numerous international conferences and events.

10.1.2. Journal

10.1.2.1. Member of the Editorial Boards

- Isabelle Puaut is Associate Editor of IEEE Transactions on Computers (IEEE TC) and Springer International Journal of Time-Critical Computing Systems.
- André Seznec is a member of the editorial boards of ACM Transactions on Architecture and Compiler Optimization.

10.1.2.2. Reviewer - Reviewing Activities

Members of PACAP routinely review submissions to numerous international journals.

10.1.3. Invited Talks

Erven Rohou and Byron Hawkins gave invited talks at the 13th *rencontres de la communauté française de compilation*.

Erven Rohou gave an invited talk at the GDR SOC2 day in Nantes.

Erven Rohou gave a talk at the Compas 2019 conference.

André Seznec was an invited speaker at the RISC V workshop in Paris.

10.1.4. Leadership within the Scientific Community

Caroline Collange is a member of the steering committee of the *Conférence francophone en informatique autour des thématiques du parallélisme, de l'architecture et des systèmes* (Compas 2019).

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).

Isabelle Puaut is member of the steering committee of the Euromicro Conference on Real Time Systems (ECRTS).

10.1.5. Research Administration

Isabelle Puaut is member of the Research Council (*Commission Recherche*) of the Université de Rennes 1. 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 Université de Rennes 1).

Erven Rohou is "correspondant scientifique des relations internationales" for Inria Rennes Bretagne Atlantique. As such he is a member of the Inria COST GTRI (Groupe de Travail "Relations Internationales").

Erven Rohou is a member of the steering committee of the high security research laboratory (LHS).

Erven Rohou is a member of the "Comité de Centre" of the Inria Rennes Research Center.

André Seznec is an elected member of the Administration Council of Inria.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

- Licence: Nicolas Bellec, Système, 14 hours, L3, Université de Rennes 1, France
- License: Nicolas Bellec, Informatique, 14 hours, L1, Lycée Chateaubriand (Rennes), France
- Arthur Blanleuil, SYS1 (Systèmes mono-utilisateur), 18 hours, M1, Université Rennes 1, France
- Arthur Blanleuil, NOY (Systèmes d'exploitation – implémentation de noyaux de systèmes), 22 hours, M1, Université Rennes 1, France
- Arthur Blanleuil, SE (Système d'Exploitation), 24 hours, M1, Université Rennes 1, France
- Master: C. Collange, GPU programming, 20 hours, M1, Université de Rennes 1, France
- Licence: D. Hardy, Real-time systems, 68 hours, L3, Université de Rennes 1, France
- Master: D. Hardy, Operating systems, 59 hours, M1, Université de Rennes 1, France
- Master: Kévin Le Bon, Architecture à Objet Canonique, 20 hours, M2, Université de Rennes 1, France
- Licence: Kévin Le Bon, SI2, 40, L1, Université de Rennes 1, France
- Master: I. Puaut, Operating systems: concepts and system programming under Linux (SEL), 77 hours, M1, Université de Rennes 1, France
- Master: I. Puaut, Operating systems kernels (NOY), 54 hours, M1, Université de Rennes 1, France
- Master: I. Puaut, Real-time systems, 40 hours, M1, Université de Rennes 1, France
- Master: I. Puaut, Optimizing and Parallelizing Compilers (OPC), 9 hours, M2, Université de Rennes 1, France

- Master: I. Puaut, Writing of scientific publications, 9 hours, M2 and PhD students, Université de Rennes 1, France
- Master: A. Seznec, Advanced Design and Architectures, 12 hours, M2 SIF, Université de Rennes 1.

10.2.2. Supervision

PhD: Arif Ali Ana-Pparakkal, *Performance Centric Dynamic Function Level Binary Transformation* [15], Université de Rennes 1, 9 Dec 2019, advisor E. Rohou

PhD in progress : Nicolas Bellec, *Security in real-time embedded systems*, started Dec 2019, advisors I. Puaut (50 %), G. Hiet from CIDRE (25 %), F. Tronel from CIDRE (25 %)

PhD in progress: Arthur Blanleuil, *Thread convergence prediction for SIMT architectures*, Université de Rennes 1, started Oct 2018, advisor C. Collange and A. Seznec

PhD in progress: Niloofar Charmchi, *Hardware prefetching and related issue*, Université de Rennes 1, started Jan 2017, advisors A. Seznec and C. Collange

PhD in progress: Kleovoulos Kalaitzidis, *Ultrawide Issue Superscalar Processors*, Université de Rennes 1, started Dec 2016, advisor A. Seznec

PhD in progress : Kévin Le Bon, *Dynamic Binary Analysis and Optimization for Cyber-Security*, started Dec 2018, advisors E. Rohou (30 %), G. Hiet from CIDRE (35 %), F. Tronel from CIDRE (35 %)

PhD in progress: Daniel Rodrigues Carvalho, *Towards a compressed memory hierarchy*, Université de Rennes 1, started Oct 2017, advisor A. Seznec

PhD in progress : Bahram Yarahmadi, *Compiler Optimizations and Worst-Case Energy Consumption*, started Feb 2018, advisor E. Rohou

10.2.3. Juries

Isabelle Puaut was a member of the following PhD thesis and habilitation thesis committees:

- Muhammad Refaat Sedky Soliman (PhD thesis), *Automated Compilation Framework for Scratchpad-based Real-Time Systems*, University of Waterloo, Canada, Juin 2019 (reviewer)
- Amine Naji (PhD thesis), *Timing Analysis for Time-Predictable Architectures*, Sorbonne Université, June 2019 (reviewer)
- Luca Santinelli (habilitation), *Mixed Criticality Modeling and Analysis Paradigms for Real Time Embedded Systems*, Habilitation à diriger des recherches, Institut National Polytechnique de Toulouse, May 2019 (reviewer)
- Roberto Medina (PhD thesis), *Deployment of Mixed-Criticality and Data-Driven Systems on Multi-core Architectures*, Université de Paris Saclay, January 2019

Isabelle Puaut was a member of the following hiring committees of assistant professors/professors:

- Professor, Université de Bretagne Occidentale, “methods and tools for the design of embedded systems, Systems on Chips, embedded systems and architectures”
- Professor, ENS Lyon
- Assistant professor, Université de Strasbourg, “performance, modeling and simulation”
- Assistant professor, ISAE-ENSMA Poitiers, “embedded real-time systems, data and model engineering”

Erven Rohou was a member of the following hiring committee (comités de sélection):

- Assistant professor position at Université de Grenoble on cloud and edge computing

Erven Rohou was a member of the following PhD thesis committees:

- Julien Proy, *Sécurisation systématique d'applications embarquées contre les attaques physiques*, June 2019 (reviewer).
- Nicolas Belleville, *Compilation pour l'application de contre-mesures contre les attaques par canal auxiliaire*, Université Grenoble Alpes, Nov 2019 (reviewer).
- Patryk Kiepas, *Performance analyses and code transformations for MATLAB applications*, Dec 2019 (reviewer).
- Paul Godard, *Parallélisation et passage à l'échelle durable d'une chaîne de traitement graphique pour l'impression professionnelle*, Université de Strasbourg, Dec 2019 (reviewer).

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

Caroline Collange is a member of the committee of the Gilles Kahn PhD prize of Société Informatique de France.

10.3.2. Articles and contents

- Erven Rohou gave an interview in the *Émergences* magazine on (volume 56, Mar 22nd 2019) on automatically masking sensitive information to protect against side-channel attacks.

10.3.3. Education

Nicolas Bellec and Isabelle Puaut taught *Basics of computer architecture*, a training of high school teachers as part of the opening of the new computer science option in the two final years before Baccalauréat, 10 hours.

Erven Rohou was invited to present the life of a researcher in computer science to middle school students (*Collège de Cesson-Sévigné)*

10.3.4. Interventions

Erven Rohou was present at the International Cybersecurity Forum in Lille (FIC https://www.forum-fic.com/en/home.htm). He presented a demo on compiler-generated countermeasures against side-channel attacks, developed within the context of the CHIST-ERA SECODE project.

10.3.5. Internal action

Erven Rohou presented the activities of PACAP at the SecDays event organized at Inria Rennes and IRISA.

11. 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] N. HALLOU, E. ROHOU, P. CLAUSS, A. KETTERLIN.*Dynamic Re-Vectorization of Binary Code*, in "SAMOS", July 2015, https://hal.inria.fr/hal-01155207

[3] 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

[4] D. HARDY, I. SIDERIS, N. LADAS, Y. SAZEIDES.*The performance vulnerability of architectural and non-architectural arrays to permanent faults*, in "MICRO 45", Vancouver, Canada, December 2012, https://hal.inria.fr/hal-00747488

[5] S. KALATHINGAL, S. COLLANGE, B. SWAMY, A. SEZNEC.*DITVA: Dynamic Inter-Thread Vectorization Architecture*, in "Journal of Parallel and Distributed Computing", October 2018, p. 1-32 [*DOI :* 10.1016/J.JPDC.2017.11.006], https://hal.archives-ouvertes.fr/hal-01655904

[6] P. MICHAUD.*Best-Offset Hardware Prefetching*, in "International Symposium on High-Performance Computer Architecture", Barcelona, Spain, March 2016 [*DOI :* 10.1109/HPCA.2016.7446087], https://hal.inria.fr/hal-01254863

[7] 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^{o} 3, 22 [*DOI :* 10.1145/2800787], https://hal.inria.fr/hal-01193178

[8] 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

[9] 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

[10] E. ROHOU, B. NARASIMHA SWAMY, A. SEZNEC.*Branch Prediction and the Performance of Interpreters - Don't Trust Folklore*, in "International Symposium on Code Generation and Optimization", Burlingame, United States, February 2015, https://hal.inria.fr/hal-01100647

[11] 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

[12] S. SARDASHTI, A. SEZNEC, D. A. WOOD.*Yet Another Compressed Cache: a Low Cost Yet Effective Compressed Cache*, in "ACM Transactions on Architecture and Code Optimization", September 2016, 25, https://hal.inria.fr/hal-01354248

[13] A. SEZNEC, P. MICHAUD.*A case for (partially)-tagged geometric history length branch prediction*, in "Journal of Instruction Level Parallelism", February 2006, http://www.jilp.org/vol8

[14] D. D. C. TEIXEIRA, S. COLLANGE, F. M. QUINTÃO 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

[15] A. A. ANAPPARAKKAL.*Performance Centric Dynamic Function Level Binary Transformation*, Université de Rennes 1 [UR1], December 2019, https://hal.inria.fr/tel-02394383

Articles in International Peer-Reviewed Journal

[16] L. BESNARD, P. PINTO, I. LASRI, J. BISPO, E. ROHOU, J. M. P. CARDOSO.*A framework for automatic and parameterizable memoization*, in "SoftwareX", July 2019, vol. 10, 100322 [*DOI :* 10.1016/J.SOFTX.2019.100322], https://hal.inria.fr/hal-02305415

[17] C. LIMA, J. CEZAR, G. VIEIRA LEOBAS, E. ROHOU, F. M. QUINTÃO PEREIRA.*Guided just-in-time specialization*, in "Science of Computer Programming", November 2019, vol. 185, 41 [*DOI :* 10.1016/J.SCICO.2019.102318], https://hal.inria.fr/hal-02314442

[18] V. A. NGUYEN, D. HARDY, I. PUAUT.*Cache-conscious Off-Line Real-Time Scheduling for Multi-Core Platforms: Algorithms and Implementation*, in "Real-Time Systems", 2019, p. 1-37, forthcoming [*DOI :* 10.4230/LIPICS.ECRTS.2017.14], https://hal.inria.fr/hal-02044110

[19] S. REDER, F. KEMPF, H. BUCHER, J. BECKER, P. ALEFRAGIS, N. S. VOROS, S. SKALISTIS, S. DERRIEN, I. PUAUT, O. OEY, T. STRIPF, C. FERDINAND, C. DAVID, P. ULBIG, D. MUELLER, U. DURAK.*Worst-Case Execution-Time-Aware Parallelization of Model-Based Avionics Applications*, in "Journal of Aerospace Information Systems", November 2019, vol. 16, n° 11, p. 521-533 [*DOI :* 10.2514/1.I010749], https://hal.archives-ouvertes.fr/hal-02383381

[20] C. SILVANO, G. AGOSTA, A. BARTOLINI, A. R. BECCARI, L. BENINI, L. BESNARD, J. BISPO, R. CMAR, J. M. P. CARDOSO, C. CAVAZZONI, D. CESARINI, S. CHERUBIN, F. FICARELLI, D. GADIOLI, M. GOLASOWSKI, A. LIBRI, J. MARTINOVIČ, G. PALERMO, P. PINTO, E. ROHOU, K. SLANINOVÁ, E. VITALI.*The ANTAREX domain specific language for high performance computing*, in "Microprocessors and Microsystems: Embedded Hardware Design (MICPRO)", July 2019, vol. 68, p. 58-73 [*DOI :* 10.1016/J.MICPRO.2019.05.005], https://hal.inria.fr/hal-02189586

[21] M. YUSUF, A. EL-MAHDY, E. ROHOU.*Runtime On-Stack Parallelization of Dependence-Free For-Loops in Binary Programs*, in "IEEE Letters of the Computer Society", March 2019, vol. 2, n° 1, p. 1-4 [*DOI :* 10.1109/LOCS.2019.2896559], https://hal.inria.fr/hal-02061340

[22] M. YUSUF, A. EL-MAHDY, E. ROHOU.*Towards Automatic Binary Runtime Loop De-Parallelization using On-Stack Replacement*, in "Information Processing Letters", May 2019, vol. 145, p. 53-57 [*DOI :* 10.1016/J.IPL.2019.01.009], https://hal.inria.fr/hal-02002812

International Conferences with Proceedings

[23] N. CHARMCHI, C. COLLANGE, A. SEZNEC.*Compressed cache layout aware prefetching*, in "SBAC-PAD 2019 - International Symposium on Computer Architecture and High Performance Computing", Campo Grande, MS, Brazil, October 2019, p. 1-4, https://hal.inria.fr/hal-02316773

[24] M. DARDAILLON, S. SKALISTIS, I. PUAUT, S. DERRIEN.*Reconciling Compiler Optimizations and WCET Estimation Using Iterative Compilation*, in "RTSS 2019 - 40th IEEE Real-Time Systems Symposium", Hong Kong, China, IEEE, December 2019, p. 1-13, https://hal.archives-ouvertes.fr/hal-02286164

[25] K. KALAITZIDIS, A. SEZNEC.*Value Speculation through Equality Prediction*, in "ICCD 2019 - 37th IEEE International Conference on Computer Design", Abu Dhabi, United Arab Emirates, IEEE, November 2019, p. 1-4, https://hal.archives-ouvertes.fr/hal-02383480

[26] R. MANCUSO, H. YUN, I. PUAUT.*Impact of DM-LRU on WCET: A Static Analysis Approach*, in "ECRTS 2019 - 31st Euromicro Conference on Real-Time Systems", Stuttgart, Germany, July 2019, p. 1-25 [*DOI :* 10.4230/LIPIcs.ECRTS.2019.17], https://hal.archives-ouvertes.fr/hal-02190255

[27] S. ROKICKI, E. ROHOU, S. DERRIEN.*Aggressive Memory Speculation in HW/SW Co-Designed Machines*, in "DATE 2019 - 22nd IEEE/ACM Design, Automation and Test in Europe", Florence, Italy, IEEE, March 2019, p. 332-335 [*DOI :* 10.23919/DATE.2019.8715010], https://hal.archives-ouvertes.fr/hal-01941876

[28] *Best Paper*
B. ROUXEL, S. SKALISTIS, S. DERRIEN, I. PUAUT.*Hiding Communication Delays in Contention-Free Execution for SPM-Based Multi-Core Architectures*, in "ECRTS 2019 - 31st Euromicro Conference on Real-Time Systems", Stuttgart, Germany, July 2019, p. 1-24 [*DOI :* 10.4230/LIPIcs.ECRTS.2019.25], https://hal.archives-ouvertes.fr/hal-02190271.

[29] M. SCHOEBERL, B. ROUXEL, I. PUAUT.*A Time-predictable Branch Predictor*, in "SAC 2019 - 34th ACM/SIGAPP Symposium on Applied Computing", Limassol, Cyprus, April 2019, p. 1-10 [*DOI :* 10.1145/3297280.3297337], https://hal.inria.fr/hal-01976187

[30] C. SILVANO, G. AGOSTA, A. BARTOLINI, A. R. BECCARI, L. BENINI, L. BESNARD, J. BISPO, R. CMAR, J. M. P. CARDOSO, C. CAVAZZONI, D. CESARINI, S. CHERUBIN, F. FICARELLI, D. GADIOLI, M. GOLASOWSKI, I. LASRI, A. LIBRI, J. MARTINOVIČ, G. PALERMO, P. PINTO, E. ROHOU, N. SANNA, K. SLANINOVÁ, E. VITALI.*Adaptive Optimization and Enforcement of Extra-Functional Properties in High Performance Computing: The ANTAREX Project*, in "PDP 2019 - 27th Euromicro International Conference on Parallel, Distributed and Network-Based Processing", Pavia, Italy, IEEE, February 2019, p. 116-123 [*DOI :* 10.1109/EMPDP.2019.8671584], https://hal.inria.fr/hal-02197811

[31] M. Y. SIRAICHI, V. F. D. SANTOS, C. COLLANGE, F. M. QUINTÃO PEREIRA.*Qubit allocation as a combination of subgraph isomorphism and token swapping*, in "OOPSLA", Athens, Greece, October 2019, vol. 3, p. 1-29 [*DOI :* 10.1145/3360546], https://hal.inria.fr/hal-02316820

[32] S. SKALISTIS, A. KRITIKAKOU.*Timely Fine-grained Interference-sensitive Run-time Adaptation of Time-triggered Schedules*, in "RTSS 2019 - 40th IEEE Real-Time Systems Symposium", Hong Kong, China, IEEE, December 2019, p. 1-13, https://hal.archives-ouvertes.fr/hal-02316392

Conferences without Proceedings

[33] N. CHARMCHI, C. COLLANGE.*Toward compression-aware prefetching*, in "COMPAS 2019 - Conférence d'informatique en Parallélisme, Architecture et Système", Anglet, France, June 2019, p. 1-9, https://hal.inria.fr/hal-02351461

[34] C. COLLANGE.*Ordinateurs quantiques : ouvrons la boîte*, in "COMPAS 2019 - Conférence d'informatique en Parallélisme, Architecture et Système", Anglet, France, June 2019, p. 1-9, https://hal.inria.fr/hal-02318324

[35] A. KOUYOUMDJIAN, C. COLLANGE, E. ROHOU.*Vers la reconfiguration adaptative de GPU pour chaque application*, in "COMPAS 2019 - Conférence d'informatique en Parallélisme, Architecture et Système", Anglet, France, June 2019, p. 1-6, https://hal.inria.fr/hal-02390821

[36] K. LE BON, B. HAWKINS, E. ROHOU, G. HIET, F. TRONEL.*Plateforme de protection de binaires configurable et dynamiquement adaptative*, in "RESSI 2019 - Rendez-Vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information", Erquy, France, May 2019, p. 1-3, https://hal.inria.fr/hal-02385216

Research Reports

[37] P. MICHAUD.*A Simple Model of Processor Temperature for Deterministic Turbo Clock Frequency*, Inria Rennes, December 2019, n° RR-9308, https://hal.inria.fr/hal-02391970

Other Publications

[38] S. ROKICKI, E. ROHOU, S. DERRIEN.*Hybrid-DBT: Hardware Accelerated Dynamic Binary Translation*, June 2019, 1, RISC-V 2019 - Workshop Zurich, Poster, https://hal.archives-ouvertes.fr/hal-02155019

References in notes

[39] G. BERTHOU, A. CARER, H.-P. CHARLES, S. DERRIEN, K. MARQUET, I. MIRO-PANADES, D. PALA, I. PUAUT, F. RASTELLO, T. RISSET, E. ROHOU, G. SALAGNAC, O. SENTIEYS, B. YARAHMADI.*The Inria ZEP project: NVRAM and Harvesting for Zero Power Computations*, March 2018, 1, NVMW 2018 - 10th Annual Non-Volatile Memories Workshop, Poster, https://hal.inria.fr/hal-01941766

[40] A. COHEN, E. ROHOU.*Processor Virtualization and Split Compilation for Heterogeneous Multicore Embedded Systems*, in "DAC", June 2010, p. 102–107

[41] 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

[42] 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

[43] R. KUMAR, D. M. TULLSEN, N. P. JOUPPI, P. RANGANATHAN.*Heterogeneous chip multiprocessors*, in "IEEE Computer", nov. 2005, vol. 38, n° 11, p. 32–38

[44] 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.archives-ouvertes.fr/hal-01087719

[45] 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

[46] 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

[47] 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

[48] 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

[49] 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^{o} 4, p. 334–346

[50] 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

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

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. - Signal processing
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
A9.2. - Machine learning
A9.3. - Signal analysis

Other Research Topics and Application Domains:

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
B9.8. - Reproducibility
B9.11.1. - Environmental risks

1. Team, Visitors, External Collaborators

Research Scientists

Frédéric Bimbot [Team leader, CNRS, Senior Researcher, HDR]
Rémi Gribonval [Team leader, Inria, Senior Researcher, until Jul 2019, HDR]
Nancy Bertin [CNRS, Researcher]
Jeremy Cohen [CNRS, Researcher]

Faculty Members

Nicolas Courty [Univ de Bretagne sud, Professor, until Aug 2019, HDR]
Lilian Besson [École Normale Supérieure Rennes, AGPR, from Nov 2019]

Technical Staff

Ewen Camberlein [Inria, Engineer, until Aug 2019]
Clément Gaultier [Inria, Engineer, from Feb 2019 until Aug 2019]
Hakim Hadj Djilani [Inria, Engineer, until Apr 2019]
Romain Lebarbenchon [Inria, Engineer, until Jul 2019]

PhD Students

Antoine Chatalic [Univ de Rennes I, PhD Student]
Diego Di Carlo [Inria, PhD Student]
Cassio Fraga Dantas [Inria, PhD Student]
Clément Gaultier [Inria, PhD Student, until Jan 2019]
Valentin Gillot [INSA Rennes, PhD Student]
Mohammed Hafsati [Institut de recherche technologique B-com, PhD Student]
Nathan Libermann [Inria, PhD Student]
Axel Marmoret [Univ de Rennes I, PhD Student, from Oct 2019]
Pierre Stock [Facebook, PhD Student, granted by CIFRE]

Post-Doctoral Fellow

Clément Elvira [Inria, Post-Doctoral Fellow, until Nov 2019]

Visiting Scientist

Andreas Loukas [EPFL, until Feb 2019]

Administrative Assistant

Stéphanie Lemaile [Inria, Administrative Assistant]

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 audiovisual streams, musical contents, biomedical data, remote sensing ...).
- 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, audio and visual 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 and remote sensing. 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.

One of the objectives of PANAMA is to instantiate and validate specific instances of audio source separation approaches and to target them 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.

4.4. Remote Sensing

Remote sensing is an image acquisition paradigm in which various pieces of information from the surface of the earth are obtained by airborne or satellite imaging. In the recent years the quality and variety of sensors available on the market for such remote acquisitions has drastically improved, leaving a large number of unsolved or difficult data processing tasks on the hand of data scientists. In particular, members of the PANAMA project-team have focused on processing hyperspectral images, that are images collected at a large number of wavelengths. The tasks that have been addressed are mainly spectral unmixing, that aims at identifying the various materials on the earth surface with their relative abundances, and denoising, that aims at removing structured noise from remote acquisitions that can be polluted by e.g. sensor movement or clouds. The spectra measured at each pixel of an hyperspectral image are heavily dependent on surface features (material, slop, material mixture type) which makes these data extremely rich and interesting.

5. Highlights of the Year

5.1. Highlights of the Year

- The **Premier Prix de Thèse de la Fondation Rennes 1** in the area of *Mathématiques, Sciences et Technologies de l'Information et de la Communication*, was awarded to **Himalaya Jain** for his Ph.D. [73] titled "Learning compact representations for large scale image search", conducted under the joint supervision of R. Gribonval and Patrick Perez, Technicolor R & I, Rennes.
- The **Prix Jeune Chercheur** from the *Journée Science et Musique 2019 (Rennes)* was awarded to **Corentin Louboutin** for a contribution titled "Modélisation multi-échelle et multi-dimensionnelle de la structure musicale", in relation to his PhD thesis [13].

6. New Software and Platforms

6.1. 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: A Software Development Initiative (ADT REVELATION) started in April 2018 for the maturation of FAuST. A first step achieved this year was to complete and robustify Matlab wrappers, to code Python wrappers with the same functionality, and to setup a continuous integration process. A second step was to simplify the parameterization of the main algorithms. The roadmap for next year includes showcasing examples and optimizing computational efficiency. – 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 multi-layer sparse approximations", IEEE Transactions on Signal and Information Processing over Networks,2017.

- Participants: Luc Le Magoarou, Nicolas Tremblay, Rémi Gribonval, Nicolas Bellot, Adrien Leman and Hakim Hadj-Djilani
- 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 Multi-layer 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/

6.2. 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 non-convex 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 2018, the code has been robustified with the help of InriaTech for a potential industrial transfer. – 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, Alexandre Sanchez, 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/

6.3. 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 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/ .

Version 2.2.2 (current version) has been released in May 2018. This version was developed in the PANAMA team through the Inria funded ADT "FFWD" (FASST For Wider Dissemination). A version 3.0 is currently under development and will be released in 2019.

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.

NEWS OF THE YEAR: Version 2.2.2 (current version) has been released in May 2018. This version was developed in the PANAMA team through the Inria funded ADT FFWD (FASST For Wider Dissemination). A version 3.0 is currently under development and will be released in 2019.

- 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/

6.4. Multi-channel BSS Locate Basic

KEYWORDS: Audio - Localization - Signal processing - Multichannel signal

SCIENTIFIC DESCRIPTION: Multi-Channel BSS Locate is a Matlab toolbox to estimate Direction Of Arrival (expressed both in azimuth and elevation) of multiple sources in a multi-channel audio signal recorded by an array of microphones. This toolbox implements the previous 8 angular spectrum methods presented in BSS Locate (GCC-PHAT, GCC-NONLIN, MUSIC and several SNR-based spectra).

NEWS OF THE YEAR: In 2018, with the help of InriaTech, selected parts of Multi-channel BSS Locate were ported to C++ in the perspective of a transfer

- Authors: Charles Blandin, Ewen Camberlein, Romain Lebarbenchon, Emmanuel Vincent, Alexey Ozerov and Nancy Bertin
- Contact: Emmanuel Vincent
- URL: http://bass-db.gforge.inria.fr/bss_locate/

6.5. VoiceHome-2

KEYWORDS: Speech processing - Audio signal processing - Source Separation - Source localization

SCIENTIFIC DESCRIPTION: New, extended version of the voiceHome corpus for distant-microphone speech processing in domestic environments. This 5-hour corpus includes short reverberated, noisy utterances (smart home commands) spoken in French by 12 native French talkers in diverse realistic acoustic conditions and recorded by an 8-microphone device at various angles and distances and in various noise conditions. Noise-only segments before and after each utterance are included in the recordings. Clean speech and spontaneous speech recorded in 12 real rooms distributed in 4 different homes are also available. All data have been fully annotated.

- Participants: Nancy Bertin, Ewen Camberlein, Romain Lebarbenchon, Emmanuel Vincent, Sunit Sivasankaran, Irina Illina and Frédéric Bimbot
- Contact: Nancy Bertin
- Publication: VoiceHome-2, an extended corpus for multichannel speech processing in real homes

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's activity ranges from theoretical results to algorithmic design and software contributions in the fields of sparse representations, inverse problems and dimension reduction.

7.1.1. Computational Representation Learning: Algorithms and Theory

Participants: Rémi Gribonval, Hakim Hadj Djilani, Cássio Fraga Dantas, Jeremy Cohen.

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. In our work introduced below, by imposing structural constraints on the dictionary and pruning provably unused atoms, we could alleviate the curse of dimensionality.

Multilayer sparse matrix products for faster computations. Inspired by usual fast transforms, we proposed a general dictionary structure (called FAμST for Flexible Approximate Multilayer Sparse Transforms) that allows cheaper manipulation, and an algorithm to learn such dictionaries together with their fast implementation, with reduced sample complexity. A comprehensive journal paper was published in 2016 [75], and we further explored the application of this technique to obtain fast approximations of Graph Fourier Transforms [76], empirically showing that $\mathcal{O}(n \log n)$ approximate implementations of Graph Fourier Transforms are possible for certain families of graphs. This opened the way to substantial accelerations for Fourier Transforms on large graphs. This year we focused on the development of the FAμST software library (see Section 6), providing transparent interfaces of FAμST data-structures with both Matlab and Python.

Kronecker product structure for faster computations. In parallel to the development of FAuST, we proposed another approach to structured dictionary learning that also aims at speeding up both sparse coding and dictionary learning. We used the fact that for tensor data, a natural set of linear operators are those that operate on each dimension separately, which correspond to rank-one multilinear operators. These rank-one operators may be cast as the Kronecker product of several small matrices. Such operators require less memory and are computationally attractive, in particular for performing efficient matrix-matrix and matrix-vector operations. In our proposed approach, dictionaries are constrained to belong to the set of low-rank multilinear operators, that consist of the sum of a few rank-one operators. The general approach, coined HOSUKRO for High Order Sum of Kronecker products, was shown last year to reduce empirically the sample complexity of dictionary learning, as well as theoretical complexity of both the learning and the sparse coding operations [67]. This year we demonstrated its potential for hyperspectral image denoising. A new efficient algorithm with lighter sample complexity requirements and computational burden was proposed and shown to be competitive with the state-of-the-art for hyperspectral image denoising with dedicated adjustements [50], [28], [27].

Combining faster matrix-vector products with screening techniques. We combined accelerated matrix-vector multiplications offered by FAμST / HOSUKRO matrix approximations with dynamic screening [59], that safely eliminates inactive variables to speedup iterative convex sparse recovery algorithms. First, we showed how to obtain safe screening rules for the exact problem while manipulating an approximate dictionary [68]. We then adapted an existing screening rule to this new framework and define a general procedure to leverage the advantages of both strategies. This lead to a journal publication [21] that includes new techniques based on duality gaps to optimally switch from a coarse dictionary approximation to a finer one. Significant complexity reductions were obtained in comparison to screening rules alone.

7.1.2. 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 [64], 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 $\ell^p \to \ell^q$ norms.

An interesting representative is the sparse pseudoinverse which we studied in much more detail in a second part of this work published this year [19], 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 $1 \leq p \leq 2$. 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, Antoine Chatalic.

Main collaborations this year: Nicolas Keriven (ENS Paris), Phil Schniter & Evan Byrne (Ohio State University, USA), Laurent Jacques & Vincent Schellekens (Univ Louvain, Belgium), Florimond Houssiau & Y.-A. de Montjoye (Imperial College London, UK)

Sketching for Large-Scale Learning. When learning from voluminous data, memory and computational time can become prohibitive. We proposed during the Ph.D. thesis of Anthony Bourrier [60] and Nicolas Keriven [74] an approach based on sketching. A low-dimensional sketch is computed by averaging (random) features over the training collection. The sketch can be seen as made of a collection of empirical generalized moments of the underlying probability distribution. Leveraging analogies with compressive sensing, we experimentally showed that it is possible to precisely estimate the mixture parameters provided that the sketch is large enough, and released an associated toolbox for reproducible research (see SketchMLBox, Section 6) with the so-called Compressive Learning Orthogonal Matching Pursuit (CL-OMP) algorithm which is inspired by Matching Pursuit. Three unsupervised learning settings have been addressed so far: Gaussian Mixture Modeling, k-means clustering, and principal component analysis. A survey conference paper on sketching for large-scale learning was published this year [25], and an extended journal version of this survey is in preparation.

Efficient algorithms to learn for sketches Last year, we showed that in the high-dimensional setting one can substantially speedup both the sketching stage and the learning stage with CL-OMP by replacing Gaussian random matrices with fast random linear transforms in the sketching procedure [63]. We studied an alternative to CL-OMP for cluster recovery from a sketch, which is 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 [61]. During his first year of Ph.D., Antoine Chatalic visited the group of Phil Schniter to further investigate this topic, and a journal paper has been published as a result of this collaboration [15].

Privacy-preserving sketches Sketching provides a potentially privacy-preserving data analysis tool, since the sketch does not explicitly disclose information about individual datum. We established theoretical privacy guarantees (with the *differential privacy* framework) and explored the utility / privacy tradeoffs of Compressive K-means [24]. A journal paper is in preparation where we extend these results to Gaussian mixture modeling and principal component analysis.

Advances in optical-based random projections Random projections are a key ingredient of sketching. Motivated by the recent development of dedicated optics-based hardware for rapid random projections, which leverages the propagation of light in random media, we tackled the problem of recovering the phase of complex linear measurements when only magnitude information is available and we control the input. A signal of interest $\xi \in \mathbb{R}^N$ is mixed by a random scattering medium to compute the projection $y = \mathbf{A}\xi$, with $\mathbf{A} \in \mathbb{C}^{M\times N}$ a realization of a standard complex Gaussian independent and identically distributed (iid) random matrix. Such optics-based matrix multiplications can be much faster and energy-efficient than their CPU or GPU counterparts, yet two difficulties must be resolved: only the intensity $|y|^2$ can be recorded by the camera, and the transmission matrix $\mathbf{A}$ is unknown. We showed that even without knowing $\mathbf{A}$, we can recover the unknown phase of y for some equivalent transmission matrix with the same distribution as $\mathbf{A}$. Our method is based on two observations: first, conjugating or changing the phase of any row of $\mathbf{A}$ does not change its distribution; and second, since we control the input we can interfere ξ with arbitrary reference signals. We showed how to leverage these observations to cast the measurement phase retrieval problem as a Euclidean distance geometry problem. We demonstrated appealing properties of the proposed algorithm in both numerical simulations and real hardware experiments. Not only does our algorithm accurately recover the missing phase, but it mitigates the effects of quantization and the sensitivity threshold, thus improving the measured magnitudes [33].

7.1.4. Theoretical results on Low-dimensional Representations, Inverse problems, and Dimension Reduction

Participants: Rémi Gribonval, Clément Elvira, Jérémy Cohen.

Main collaboration: Nicolas Keriven (ENS Paris), Gilles Blanchard (Univ Postdam, Germany), Cédric Herzet (SIMSMART project-team, IRMAR / Inria Rennes), Charles Soussen (Centrale Supelec, Gif-sur-Yvette), Mila Nikolova (CMLA, Cachan), Nicolas Gillis (UMONS)

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 is being revised for a journal [71].

Recovery guarantees for algorithms with continuous dictionaries. We established theoretical guarantees on sparse recovery guarantees for a greedy algorithm, orthogonal matching pursuit (OMP), in the context of continuous dictionaries [66], e.g. as appearing in the context of sparse spike deconvolution. Analyses based on discretized dictionary fail to be conclusive when the discretization step tends to zero, as the coherence goes to one. Instead, our analysis is directly conducted in the continuous setting amd exploits specific properties of the positive definite kernel between atom parameters defined by the inner product between the corresponding atoms. For the Laplacian kernel in dimension one, we showed in the noise-free setting that OMP exactly recovers the atom parameters as well as their amplitudes, regardless of the number of distinct atoms [66]. A preprint describing a full class of kernels for which such an analysis holds, in particular for higher dimensional parameters, has been released and submitted to a journal [30], [36], [31], [51].

Identifiability of Complete Dictionary Learning In the era of deep learning, dictionary learning has proven to remain an important and extensively-used data mining and processing tool. Having been studied and used for over twenty years, dictionary learning has well-understood properties. However there was a particular stone missing, which was understanding deterministic conditions for the parameters of dictionary learning to be uniquely retrieved from a training data set. We filled this gap partially by drastically improving on the previously best such conditions in the case of complete dictionaries [16]. Moreover, although algorithms with guaranties to compute the unique best solution do exist, they are seldom used in practice due to their high computational cost. In subsequent work, we showed that faster algorithms typically used to compute dictionary learning often failed at computing the unique solution (in cases where our previous result guaranties this uniqueness), opening the way to new algorithms that are both fast and guarantied [26].

On Bayesian estimation and proximity operators. There are two major routes to address the ubiquitous family of inverse problems appearing in signal and image processing, such as denoising or deblurring. The first route is Bayesian modeling: prior probabilities are used to model both the distribution of the unknown variables and their statistical dependence with the observed data, and estimation is expressed as the minimization of an expected loss (e.g. minimum mean squared error, or MMSE). The other route is the variational approach, popularized with sparse regularization and compressive sensing. It consists in designing (often convex) optimization problems involving the sum of a data fidelity term and a penalty term promoting certain types of unknowns (e.g., sparsity, promoted through an L1 norm).

Well known relations between these two approaches have lead to some widely spread misconceptions. In particular, while the so-called Maximum A Posterori (MAP) estimate with a Gaussian noise model does lead to an optimization problem with a quadratic data-fidelity term, we disprove through explicit examples the common belief that the converse would be true. In previous work we showed that for denoising in the presence of additive Gaussian noise, for any prior probability on the unknowns, the MMSE is the solution of a penalized least-squares problem, with all the apparent characteristics of a MAP estimation problem with Gaussian noise and a (generally) different prior on the unknowns [72]. In other words, the variational approach is rich enough to build any MMSE estimator associated to additive Gaussian noise via a well chosen penalty.

This year, we achieved generalizations of these results beyond Gaussian denoising and characterized noise models for which the same phenomenon occurs. In particular, we proved that with (a variant of) Poisson noise and any prior probability on the unknowns, MMSE estimation can again be expressed as the solution of a penalized least-squares optimization problem. For additive scalar denoising, the phenomenon holds if and only if the noise distribution is log-concave, resulting in the perhaps surprising fact that scalar Laplacian denoising can be expressed as the solution of a penalized least-squares problem [22] Somewhere in the proofs appears an apparently new characterization of proximity operators of (nonconvex) penalties as subdifferentials of convex potentials [54].

7.1.5. Low-rank approximations: fast constrained algorithms

Participant: Jeremy Cohen.

Main collaborations: Nicolas Gillis (Univ. Mons, Belgium), Andersen Man Shun Ang (Univ. Mons, Belgium), Nicolas Nadisic (Univ. Mons, Belgium).

Low-Rank Approximations (LRA) aim at expressing the content of a multiway array by a sum of simpler separable arrays. Understood as a powerful unsupervised machine learning technique, LRA are most and foremost modern avatars of sparsity that are still not fully understood. In particular, algorithms to compute the parameters of LRA demend a lot of computer ressources and provide sub-optimal results. An important line of work over the last year has been to design efficient algorithms to compute constrained LRA, and in particular constrained low-rank tensor decompositions. This work has been carried out though a collaboration with the ERC project COLORAMAP of Nicolas Gillis (Univ. Mons, Belgium) and his PhD students Nicolas Nadisic (co-supervision) and Andersen Man Shun Ang.

Extrapolated Block-coordinate algorithms for fast tensor decompositions State-of-the-art algorithms for computing tensor decompositions are based on the idea that solving alternatively for smaller blocks of parameters is easier than solving the large problem at once. Despite showing nice convergence speeds, the obtained Block Coordinate Descent algorithms (BCD) are prone to being stuck near saddle points. We have shown in preliminary work, which is still ongoing, that BCD algorithms can be improved using Nesterov extrapolation in-between block updates. This improves empirical convergence speed in constrained and unconstrained tensor decompositions tremendously at almost no additional computation cost, and is therefore bound to have a large impact on the community [37].

Exact sparse nonnegative least-squares solutions to least-squares problems Another important LRA is Nonnegative Matrix factorization, which has found many diverse applications such as in remote sensing or automatic music transcription. Sometimes, imposing sparsity on parameters of NMF is crucial to be able to correctly process and interpret the output of NMF. However, sparse NMF has scarcely been studied, and its computation is challenging. In fact, even only a subproblem in a BCD approach, sparse nonnegative least-squares, is already NP-hard. We proposed to solve this sparse nonnegative least-squares problem exactly using a combinatorial algorithm. To reduce as much as possible the cost of solving this combinatorial problem, a Branch and Bound algorithm was proposed which, on average, reduces the computational complexity drastically. A next step will be to use this branch and bound algorithm as a brick for proposing an efficient algorithm for sparse NMF.

7.1.6. Algorithmic Exploration of Sparse Representations for Neurofeedback

Participant: Rémi Gribonval.

Claire Cury, Pierre Maurel & Christian Barillot (EMPENN Inria project-team, Rennes)

In the context of the HEMISFER (Hybrid Eeg-MrI and Simultaneous neuro-feedback for brain Rehabilitation) Comin Labs project (see Section 1), in collaboration with the EMPENN team, we validated a technique to estimate brain neuronal activity by combining EEG and fMRI modalities in a joint framework exploiting sparsity [82]. We then focused on directly estimating neuro-feedback scores rather than brain activity. Electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) both allow measurement of brain activity for neuro-feedback (NF), respectively with high temporal resolution for EEG and high spatial resolution for fMRI. Using simultaneously fMRI and EEG for NF training is very promising to devise brain rehabilitation protocols, however performing NF-fMRI is costly, exhausting and time consuming, and cannot be repeated too many times for the same subject. We proposed a technique to predict NF scores from EEG recordings only, using a training phase where both EEG and fMRI NF are available [39]. A journal paper has been submitted.

7.2. Emerging activities on high-dimensional learning with neural networks

Participants: Rémi Gribonval, Himalaya Jain, Pierre Stock.

Main collaborations: Patrick Perez (Technicolor R & I, Rennes), Gitta Kutyniok (TU Berlin, Germany), Morten Nielsen (Aalborg University, Denmark), Felix Voigtlaender (KU Eichstätt, Germany), Herve Jegou and Benjamin Graham (FAIR, Paris)

dictionary learning, large-scale indexing, sparse deep networks, normalization, sinkhorn, regularization

Many of the data analysis and processing pipelines that have been carefully engineered by generations of mathematicians and practitioners can in fact be implemented as deep networks. Allowing the parameters of these networks to be automatically trained (or even randomized) allows to revisit certain classical constructions. Our team has started investigating the potential of such approaches both from an empirical perspective and from the point of view of approximation theory.

Learning compact representations for large-scale image search. The PhD thesis of Himalaya Jain [73], which received the Fondation Rennes 1 PhD prize this year, was dedicated to learning techniques for the design of new efficient methods for large-scale image search and indexing.

Equi-normalization of Neural Networks. Modern neural networks are over-parameterized. In particular, each rectified linear hidden unit can be modified by a multiplicative factor by adjusting input and output weights, without changing the rest of the network. Inspired by the Sinkhorn-Knopp algorithm, we introduced a fast iterative method for minimizing the l2 norm of the weights, equivalently the weight decay regularizer. It provably converges to a unique solution. Interleaving our algorithm with SGD during training improves the test accuracy. For small batches, our approach offers an alternative to batch- and group- normalization on CIFAR-10 and ImageNet with a ResNet-18. This work was presented at ICLR 2019 [41].

Approximation theory with deep networks. We study the expressivity of sparsely connected deep networks. Measuring a network's complexity by its number of connections with nonzero weights, or its number of neurons, we consider the class of functions which error of best approximation with networks of a given complexity decays at a certain rate. Using classical approximation theory, we showed that this class can be endowed with a norm that makes it a nice function space, called approximation space. We established that the presence of certain "skip connections" has no impact of the approximation space, and studied the role of the network's nonlinearity (also known as activation function) on the resulting spaces, as well as the benefits of depth. For the popular ReLU nonlinearity (as well as its powers), we related the newly identified spaces to classical Besov spaces, which have a long history as image models associated to sparse wavelet decompositions. The sharp embeddings that we established highlight how depth enables sparsely connected networks to approximate functions of increased "roughness" (decreased Besov smoothness) compared to shallow networks and wavelets. A preprint has been published and is under review for a journal [23].

7.3. Emerging activities on Nonlinear Inverse Problems

Compressive sensing, compressive learning, audio inpainting, phase estimation

7.3.1. Audio Inpainting and Denoising

Participants: Rémi Gribonval, Nancy Bertin, Clément Gaultier.

Main collaborations: Srdan Kitic (Orange, 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 [57], we proposed over the last five years a series of algorithms leveraging the competitive cosparse approach, which offers a very appealing trade-off between reconstruction performance and computational time, and its extensions to the incorporation of the so-called "social" into problems regularized by a cosparse prior. We exhibited a common framework allowing to tackle both denoising and declipping in a unified fashion [69]; these results, together with listening tests results that were specified and prepared in 2019 and will be run soon, will be included in an ongoing journal paper, to be submitted in 2020. This year, following Clément Gaultier Ph.D. defense [12], we progressed towards industrial transfer of these results through informal interaction with a company commercializing audio plugins, in particular with new developments to alleviate some artifacts absent from simulation but arising in real-world use cases.

7.4. Source Localization and Separation

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 [84] 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 [11].

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 INVATE project and the collaboration with the startup 5$^{\text{th}}$ dimension (see Sections 8.1.2, 8.1.4), on the one hand; on the other hand, an emerging track on audio scene analysis with machine learning, evolved beyond the "localization and separation" paradigm, and is the subject of a more recent axis of research presented in Section 7.5.

7.4.1. Towards Real-world Localization and Separation

Participants: Nancy Bertin, Frédéric Bimbot, Rémi Gribonval, Ewen Camberlein, Romain Lebarbenchon, Mohammed Hafsati.

Main collaborations: Emmanuel Vincent (MULTISPEECH Inria project-team, Nancy)

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, our work within the voiceHome project (2015-2017), an OSEO-FUI industrial collaboration [0] 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) found its conclusion with the publication of a journal paper in a special issue of Speech Communication [14].

Accomplished progress and levers of improvements identified thanks to this project resulted in the granting of an Inria ADT (Action de Développement Technologique). This new development phase of the FASST software started in September 2017 and was achieved this year by the release of the third version of the toolbox, with significant progress towards efficient initialization, low latency and reduction of the computational burden.

In addition, evolutions of the MBSSLocate software initiated during this project led to a successful participation in the IEEE-AASP Challenge on Acoustic Source Localization and Tracking (LOCATA) [77], and served as a baseline for the publication of the for the IEEE Signal Processing Cup 2019 [21]. The SP Cup was also fueled by the publicly available DREGON dataset 5 recorded in PANAMA, including noiseless speech and on-flight ego-noise recordings, devoted to source localization from a drone [117].

Finally, these progress also led to a new industrial transfer with the start-up 5$^{\text{th}}$ dimension (see Section 8.1.4). During this collaboration aiming at equipping a pair of glasses with an array of microphones and "smart" speech enhancement functionalities, we particularly investigated the impact of obstacles between microphones in the localization and separation performance, the selection of the best subset of microphones in the array for side speakers hidden by the head shadow, and the importance of speaker enrolment (learning spectral dictionaries of target users voices) in this use case.

7.4.2. Separation for Remixing Applications

Participants: Nancy Bertin, Rémi Gribonval, Mohammed Hafsati.

Main collaborations: Nicolas Epain (IRT b<>com, Rennes)

Second, through the Ph.D. of Mohammed Hafsati (in collaboration with the IRT b<>com with the INVATE project, see Section 8.1.2) started in November 2016, we investigated a new application of source separation to sound re-spatialization from Higher Order Ambisonics (HOA) signals [70], 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. Simulation results showed that separating sources in the HOA domain results in a 5 to 15 dB increase in signal-to-distortion ratio, compared to the microphone domain. These results were accepted for publication in the DAFx international conference [34]. We continued extending our methods following two tracks: hybrid acquisition scenarios, where the separation

[0] With partners: onMobile, Delta Dore, eSoftThings, Orange, Technicolor, LOUSTIC, Inria Nancy.

of HOA signals can be informed by complementary close-up microphonic signals, and the replacement of spectrogram NMF by neural networks for a better spectral adaptation of the models. Future work will include subjective evaluation of the developed workflows.

7.5. Towards comprehensive audio scene analysis

Source localization and separation, machine learning, room geometry, room properties, multichannel audio classification

By contrast to the previous lines of work and results on source localization and separation, which are mostly focused on the *sources*, the following emerging activities consider the audio scene and its analysis in a wider sense, including the environment around the sources, and in particular the *room* they are included in, and their properties. This inclusive vision of the audio scene allows in return to revisit classical audio processing tasks, such as localization, separation or classification.

7.5.1. Room Properties: Estimating or Learning Early Echoes

Participants: Nancy Bertin, Diego Di Carlo, Clément Elvira.

Main collaborations: Antoine Deleforge (Inria Nancy – Grand Est), Ivan Dokmanic (University of Illinois at Urbana-Champaign, Coordinated Science Lab, USA), Robin Scheibler (Tokyo Metropolitan University, Tokyo, Japan), Helena Peic-Tukuljac (EPFL, Switzerland).

In [85] we showed that the knowledge of early echoes improved sound source separation performances, which motivates the development of (blind) echo estimation techniques. Echoes are also known to potentially be a key to the room geometry problem [65]. In 2019, two different approaches to this problem were explored.

As a competitive, yet similar approach to our previous work in [83], we proposed a new analytical method for off-the-grid early echoes estimation, based on continuous dictionaries and extensions of sparse recovery methods in this setting. From the well-known *cross-relation* between room impulse responses and signals in a "one source - two microphones" settings, the echo estimation problem can be recast as a Beurling-LASSO problem and solved with algorithms of this kind. This enables near-exact blind and off-grid echo retrieval from discrete-time measurements, and can outperform conventional methods by several orders of magnitude in precision, in an ideal case where the room impulse response is limited to a few weighted Diracs. Future work will include alternative initialization schemes, extensions to sparse-spectrum signals and noisy measurements, and applications to dereverberation and audio-based room shape reconstruction. This work, mostly lead by Clément Elvira, was submitted for publication in Icassp 2020.

On the other hand, the PhD thesis of Diego Di Carlo aims at applying the "Virtual Acoustic Space Traveler" (VAST) framework to the blind estimation of acoustic echoes, or other room properties (such as reverberation time, acoustic properties at the boundaries, etc.) Last year, we focused on identifying promising couples of inputs and outputs for such an approach, especially by leveraging the notions of relative transfer functions between microphones, the room impulse responses, the time-difference-of-arrivals, the angular spectra, and all their mutual relationships. In a simple yet common scenario of 2 microphones close to a reflective surface and one source (which may occur, for instance, when the sensors are placed on a table such as in voice-based assistant devices), we introduced the concept of microphone array augmentation with echoes (MIRAGE) and showed how estimation of early-echo characteristics with a learning-based approach is not only possible but can in fact benefit source localization. In particular, it allows to retrieve 2D direction of arrivals from 2 microphones only, an impossible task in anechoic settings. These first results were published in ICASSP [29]. In 2019, we improved the involved DNN architecture in MIRAGE and worked towards experimental validation of this result, by designing and recording a data set with annotated echoes in different conditions of reverberation. Future work will include extension of this data set, extension to more realistic and more complex scenarios (including more microphones, sources and reflective surfaces) and the estimation of other room properties such as the acoustic absorption at the boundaries, or ultimately, the room geometry. Some of these tracks currently benefit from the visit of Diego di Carlo to Bar-Ilan University (thanks to a MathSTIC doctoral outgoing mobility grant.)

7.5.2. Multichannel Audio Event and Room Classification

Participants: Marie-Anne Lacroix, Nancy Bertin.

Main collaborations: Pascal Scalart, Romuald Rocher (GRANIT Inria project-team, Lannion)

Typically, audio event detection and classification is tackled as a "pure" single-channel signal processing task. By constrast, audio source localization is the perfect example of multi-channel task "by construction". In parallel, the need to classify the type of scene or room has emerged, in particular from the rapid development of wearables, the "Internet of things" and their applications. The PhD of Marie-Anne Lacroix, started in September 2018, combines these ideas with the aim of developing multi-channel, room-aware or spatially-aware audio classification algorithms for embedded devices. The PhD topic includes low-complexity and low-energy stakes, which will be more specifically tackled thanks to the GRANIT members area of expertise. During the first year of the PhD, we gathered existing data and identified the need for new simulations or recordings, and combined ideas from existing single-channel classification techniques with traditional spatial features in order to design several baseline algorithms for multi-channel joint localization and classification of audio events. The impact of feature quantization on classification performance is also currently under investigation and a participation to the 2020 edition of the IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events (DCASE) is envisioned.

7.6. 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 [58], in particular the System & Contrast model [1].

7.6.1. Modeling music by Polytopic Graphs of Latent Relations (PGLR)

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 [79] defined as the shortest displacement of notes, in semitones, between a pair of chords. Other chord representations and relations are possible, as studied in [81] 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 [80] and are related to 6 main redundant sequences which can be viewed as canonical multiscale structural patterns.

In parallel, recent work has also been dedicated to modeling melodic and rhythmic motifs in order to extend the polytopic model to multiple musical dimensions.

Results obtained in this framework illustrate the efficiency of the proposed model in capturing structural information within musical data and support the view that musical content can be delinearised in order to better describe its structure. Extensive results are included in Corentin Louboutin's PhD [13], defended in March 2019 and which was awarded the Prix Jeune Chercheur Science et Musique, in October.

7.6.2. Exploring Structural Dependencies in Melodic Sequences using Neural Networks

Participants: Nathan Libermann, Frédéric Bimbot.

This work is carried out in the framework of a PhD, co-directed by Emmanuel Vincent (Inria-Nancy).

In order to be able to generate structured melodic phrases and section, we explore various schemes for modeling dependencies between notes within melodies, using deep learning frameworks.

A a first set of experiments, we have considered a GRU-based sequential learning model, studied under different learning scenarios in order to better understand the optimal architectures in this context that can achieve satisfactory results. By this means, we wish to explore different hypotheses relating to temporal non-invariance relationships between notes within a structural segment (motif, phrase, section).

We have defined three types of recursive architectures corresponding to different ways to exploit the local history of a musical note, in terms of information encoding and generalization capabilities.

Initially conducted on the Lakh MIDI dataset, experiments have switched to the Meertens Tune Collections data set (Dutch traditional melodies) and confirm the trends observed in [78], w.r.t. the utility of non-ergodic models for the generation of melodic segments.

Ongoing work is extending these findings to the design of specific NN architectures, which incorporate attention models, to account for this non-invariance of information across musical segments.

7.6.3. Graph Signal Processing for Multiscale Representations of Music Similarity

Participants: Valentin Gillot, Frédéric Bimbot.

"Music Similarity" is a multifaceted concept at the core of Music Information Retrieval (MIR). Among the wide range of possible definitions and approaches to this notion, a popular one is the computation of a so-called content-based similarity matrix (S), in which each coefficient is a similarity measure between descriptors of short time frames at different instants within a music piece or a collection of pieces.

Matrix S can be seen as the adjacency matrix of an underlying graph, embodying the local and non-local similarities between parts of the music material. Considering the nodes of this graph as a new set of indices for the original music frames or pieces opens the door to a "delinearized" representation of music, emphasizing its structure and its semiotic content.

Graph Signal Processing (GSP) is an emerging topic devoted to extend usual signal processing tools (Fourier analysis, filtering, denoising, compression, ...) to signals "living" on graphs rather than on the time line, and to exploit mathematical and algorithmic tools on usual graphs, in order to better represent and manipulate these signals. Toy applications of GSP concepts on music content in music resequencing and music inpainting are illustrating this trend.

From exploratory experiments, first observations point towards the following hypotheses :

- local and non-local structures of a piece are highlighted in the adjacency matrix built from a simple time-frequency representation of the piece,
- the first eigenvectors of the graph Laplacian provide a rough structural segmentation of the piece,
- clusters of frames built from the eigenvectors contain similar, repetitive sound sequences.

The goal of Valentin Gillot's PhD is to consolidate these hypotheses and investigate further the topic of Graph Signal Processing for music, with more powerful conceptual tools and experiments at a larger scale.

The core of the work will consist in designing a methodology and implement an evaluation framework so as to (i) compare different descriptors and similarity measures and their capacity to capture relevant structural information in music pieces or collection of pieces, (ii) explore the structure of musical pieces by refining the frame clustering process, in particular with a multi-resolution approach, (iii) identify salient characteristics of graphs in relation to mid-level structure models and (iv) perform statistics on the typical properties of the similarity graphs on a large corpus of music in relation to music genres and/or composers.

By the end of the PhD, we expect the release of a specific toolbox for music composition, remixing and repurposing using the concepts and algorithms developed during the PhD. First results obtained this year in music recomposition have proven very conclusive [32].

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. Labex Comin Labs projects

CominLabs is a Laboratoire d'Excellence funded by the PIA (Programme Investissements d'Avenir) in the broad area of telecommunications.

- **HEMISFER (2014-2017) and HEMISFER-CLINICAL (2018-2019)**
  Participant: Rémi Gribonval.

  Acronym: HYBRID (Hybrid Eeg-MrI and Simultaneous neuro-feedback for brain Rehabilitation)

  http://hemisfer.cominlabs.u-bretagneloire.fr/

  Research axis: 3.1

  CominLabs partners : EMPENN, 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, EMPENN 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 EMPENN 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.

- **TEPN**

Participant: Rémi Gribonval.

Acronym: TEPN (Toward Energy Proportional Networks)

http://tepn.cominlabs.u-bretagneloire.fr/

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).

- **FAWI** (2019-2020)
  Fourier Adaptive Waveform Implementation
  Participant: Rémi Gribonval.
  - This project is a follow-up to TEPN. Its main goal is to implement a prototype demonstrating concretely the feasibility of the new Fourier Adaptive Waveform modulation which has been patented [62].
  - Contribution of PANAMA: to provide initial training to the recruited engineer in charge of the implementation.
  - *Partners: PANAMA, IETR.*
  - Funding: 18 months of engineer, hosted by IETR.
- **SPARSE** (2019)
  Sparse representations in continuous dictionaries
  Participants: Rémi Gribonval, Clément Elvira, Clément Merdrignac.
  - This short exploratory action aims to explore the new paradigm of sparse representations in “continuous” dictionaries.
  - Contribution of PANAMA: to design algorithms for the sparse representation problem in continuous dictionaries with theoretical success guarantees.
  - *Partners: PANAMA, SIMSMART (Inria-Rennes), ENSTA Bretagne, IMT Atlantique.*
  - Funding: 5.6kEuros (internship + travel)

8.1.2. 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.

8.1.3. ANR OATMIL project

Participants: Rémi Gribonval, Antoine Chatalic, Nicolas Courty.

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

8.1.4. Collaboration with 5th dimension – dynamic separation of localized sound sources

Participants: Nancy Bertin, Ewen Camberlein, Romain Lebarbenchon.

Duration: 1 year (2018-2019)

Research axis: 3.2

Partner: 5^{th} dimension (https://5dimo.com/)

Funding: LABEX AMIEX (https://www.agence-maths-entreprises.fr/a/)

After a first phase of this contract which involved porting in C++ a subset of our source localization library Multichannel BSS Locate (Oct.-Nov. 2018, in collaboration with InriaTech), a second phase was realized in 2019 with support from LABEX AMIES. We specified and recorded new data adapted to the partner's use case (microphones on glasses temples) and investigated the interplay between localization and separation, using the FASST library, on simulated and real data recorded with a prototype.

8.2. International Initiatives

8.2.1. Inria International Partners

8.2.1.1. Informal International Partners

Nancy Bertin is "external collaborator" of the MERLIN project (project between the Acoustics Research Institute of the Austrian Academy of Sciences and the Signal Processing Laboratory at Brno University of Technology.)

8.3. International Research Visitors

8.3.1. Visits of International Scientists

8.3.1.1. Internships

- Pavel Záviška and Ondřej Mokrý, visiting students from Brno University of Technology, in December 2018 (within the MERLIN collaboration).
- Andersen Man Shun Ang, visiting student from University of Mons, in February 2019.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Organisation

- Nancy Bertin, event coordinator of the 2019 Science and Music Day (Journée Science et Musique) organized by IRISA *(with Claire Cury from EMPENN project-team.)*
- Nancy Bertin, member of the LVA/ICA conference steering committee.

9.1.2. Scientific Events: Selection

9.1.2.1. Chair of Conference Program Committees

- Nancy Bertin, coordinator of the Science and Music Young Researcher Award (Prix Jeune Chercheur Science et Musique).
- Rémi Gribonval and Frédéric Bimbot, scientific coordinators of the Science and Music Day (Journée Science et Musique) organized by IRISA.

9.1.2.2. Member of Conference Program Committees

- Rémi Gribonval, member of the program committee of the GRETSI.

9.1.3. Membership of Editorial Boards

- Frédéric Bimbot, Editor-in-Chief of the international journal *Speech Communication*.
- Rémi Gribonval, Associate Editor of the international journal *Constructive Approximation*.
- Rémi Gribonval, Senior Area Editor of the *IEEE Transactions on Signal Processing*.
- Nancy Bertin, Guest Editor of a special issue in the international journal *IEEE Journal of Selected Topics in Signal Processing*.

9.1.4. Invited Talks

- Jérémy Cohen, invited speaker at workshop "Low Rank Optimization and Applications" (Max Plank Institute for the Mathematics in the Sciences, Leipzig, Germany, April 2019) [0]
- Jérémy Cohen, invited speaker at workshop "AI and Tensor Factorizations" (Santa Fe, USA, September 2019)
- Rémi Gribonval, keynote at SampTA 2019 (Sampling Theory and Applications) in Bordeaux, France.
- Rémi Gribonval, tutorial at Summer School "Sparsity for Physics, Signal and Learning", Inria Paris, June 2019
- Rémi Gribonval, invited talk at 4TU meeting on "Mathematics of Deep Learning", Delft University, Nov 2019
- Rémi Gribonval, invited talk at Alan Turing Institute on "Mathematics of Data", London, May 2019
- Rémi Gribonval, invited talk at GDR ISIS meeting on "Theory of Deep Learning", Paris, OCt 2019
- Rémi Gribonval, panel session on "Signal processing and AI" at GRETSI 2019, Lille, Aug 2019

[0] https://www.mis.mpg.de/calendar/conferences/2019/lroa2019.html

9.1.5. Scientific Expertise

- Rémi Gribonval, vice-president of the Scientific Advisory Board of the Acoustics Research Institute from the Austrian Academy of Sciences in Vienna.
- Frédéric Bimbot, member of the International Advisory Council of ISCA (International Speech Communication Association).
- Frédéric Bimbot, member of the ANR Expert Scientific Committee n°38, dedicated to the Digital Revolution and its relations to Knowledge and Culture.
- Rémi Gribonval, member of the EURASIP Special Area Team (SAT) on Signal and Data Analytics for Machine Learning (SiG-DML) since 2015.

9.1.6. Research Administration

- Frédéric Bimbot, board member of the GDR MADICS
- Rémi Gribonval, member of the organization committee of the 2019 GDR ISIS / GRETSI / Club EAA thesis prize in signal and image processing.
- Nancy Bertin, elected member of the “Comité de Centre” at Inria Rennes Bretagne Atlantique research center.
- Nancy Bertin, coordinator of IRISA “Arts, Heritage and Culture crosscutting axis” *(with Valérie Gouranton, HYBRID project-team.)*

9.2. Teaching

Master : N. Bertin, coordination of the VAI module "Vocal and Acoustic Interactions" within the SIF M2, 20 hours, Université de Rennes 1, 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, "Foundations of Smart Sensing", 9 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", 8 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.

Master : J. Cohen, "Parcimony", 9 hours, ENSAI / "Smart Sensing" course

Master : J. Cohen, "Sparsity in Signal and Image Processing", 19 hours, M2, Institut National des Sciences Appliquées (INSA) de Rennes, France

9.3. Popularization

9.3.1. Journée Science et Musique (JSM 2019)

Participants: Nancy Bertin, Rémi Gribonval, Nancy Bertin, Frédéric Bimbot, Ewen Camberlein, Stéphanie Gosselin-Lemaile, Corentin Louboutin, Antoine Chatalic, Clément Gaultier, Cássio Fraga Dantas, Diego Di Carlo.

with contributions and support from: Claire Cury, Elisabeth Lebret, Valérie Gouranton, Ronan Gaugne, Florian Nouviale, 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 2019 hosted approximately 300 visitors and was a partner of the “Festival des Sciences” and an official event of CNRS 80th birthday celebrations.

9.3.2. Participation to Equinoxes#1

Frédéric Bimbot gave a popularization conference at the **Equinoxes#1** event, organised by Artoutaï (13 October 2019), on "the mysterious spectrum of sound" (4 times during the day)

9.3.3. Internal or external responsibilities

Frédéric Bimbot is a member of a working group of the CNRS/DR17 on the topic "changer le regard sur le handicap au travail"

10. 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^{o} 5, p. 631-661 [*DOI :* 10.1525/mp.2016.33.5.631], https://hal.inria.fr/hal-01188244

[2] M. Chafii. *Study of a new multicarrier waveform with low PAPR*, CentraleSupélec, October 2016, https://hal.archives-ouvertes.fr/tel-01399509

[3] N. Duong, E. Vincent, R. Gribonval. *Under-determined reverberant audio source separation using a full-rank spatial covariance model*, in "IEEE Transactions on Audio, Speech and Language Processing", July 2010, vol. 18, n^{o} 7, p. 1830–1840 [*DOI :* 10.1109/TASL.2010.2050716], https://hal.inria.fr/inria-00541865

[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^{o} 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^{o} 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^{o} 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^{o} 4, p. 1118 - 1133, 16, http://hal.inria.fr/hal-00626962

[10] H. PEIC TUKULJAC, A. DELEFORGE, R. GRIBONVAL.*MULAN: A Blind and Off-Grid Method for Multichannel Echo Retrieval*, in "NIPS 2018 - Thirty-second Conference on Neural Information Processing Systems", Montréal, Canada, December 2018, p. 1-11, https://arxiv.org/abs/1810.13338 , https://hal.inria.fr/hal-01906385

[11] 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

[12] C. GAULTIER.*Design and evaluation of sparse models and algorithms for audio inverse problems*, Université Rennes 1, January 2019, https://tel.archives-ouvertes.fr/tel-02148598

[13] C. LOUBOUTIN.*Multi-scale and multi-dimensional modelling of music structure using polytopic graphs*, Université Rennes 1, March 2019, https://tel.archives-ouvertes.fr/tel-02149728

Articles in International Peer-Reviewed Journal

[14] N. BERTIN, E. CAMBERLEIN, R. LEBARBENCHON, E. VINCENT, S. SIVASANKARAN, I. ILLINA, F. BIMBOT.*VoiceHome-2, an extended corpus for multichannel speech processing in real homes*, in "Speech Communication", January 2019, vol. 106, p. 68-78 [*DOI :* 10.1016/J.SPECOM.2018.11.002], https://hal.inria.fr/hal-01923108

[15] E. BYRNE, A. CHATALIC, R. GRIBONVAL, P. SCHNITER.*Sketched Clustering via Hybrid Approximate Message Passing*, in "IEEE Transactions on Signal Processing", September 2019, vol. 67, n^{o} 17, p. 4556-4569, https://arxiv.org/abs/1712.02849 [*DOI :* 10.1109/TSP.2019.2924585], https://hal.inria.fr/hal-01991231

[16] J. E. COHEN, N. GILLIS.*Identifiability of Complete Dictionary Learning*, in "SIAM Journal on Mathematics of Data Science", 2019, vol. 1, n^{o} 3, p. 518–536, https://arxiv.org/abs/1808.08765 [*DOI :* 10.1137/18M1233339], https://hal.archives-ouvertes.fr/hal-02183578

[17] C. CURY, P. MAUREL, R. GRIBONVAL, C. BARILLOT.*A sparse EEG-informed fMRI model for hybrid EEG-fMRI neurofeedback prediction*, in "Frontiers in Neuroscience", January 2020 [*DOI :* 10.3389/FNINS.2019.01451], https://www.hal.inserm.fr/inserm-02090676

[18] A. DELEFORGE, D. DI CARLO, M. STRAUSS, R. SERIZEL, L. MARCENARO.*Audio-Based Search and Rescue with a Drone: Highlights from the IEEE Signal Processing Cup 2019 Student Competition*, in "IEEE Signal Processing Magazine", September 2019, vol. 36, n^{o} 5, p. 138-144, https://arxiv.org/abs/1907.04655 [*DOI :* 10.1109/MSP.2019.2924687], https://hal.archives-ouvertes.fr/hal-02161897

[19] I. DOKMANIĆ, R. GRIBONVAL.*Concentration of the Frobenius norm of generalized matrix inverses*, in "SIAM Journal on Matrix Analysis and Applications", 2019, vol. 40, n^o 1, p. 92–121, https://arxiv.org/abs/1810.07921 - Revised/condensed/renamed version of preprint "Beyond Moore-Penrose Part II: The Sparse Pseudoinverse", forthcoming [*DOI :* 10.1137/17M1145409], https://hal.inria.fr/hal-01897046

[20] S. FOUCART, R. GRIBONVAL, L. JACQUES, H. RAUHUT.*Jointly Low-Rank and Bisparse Recovery: Questions and Partial Answers*, in "Analysis and Applications", 2020, vol. 18, n^o 01, p. 25–48, https://arxiv.org/abs/1902.04731 [*DOI :* 10.1142/S0219530519410094], https://hal.inria.fr/hal-02062891

[21] C. FRAGA DANTAS, R. GRIBONVAL.*Stable safe screening and structured dictionaries for faster L1 regularization*, in "IEEE Transactions on Signal Processing", July 2019, vol. 67, n^o 14, p. 3756-3769, https://arxiv.org/abs/1812.06635 [*DOI :* 10.1109/TSP.2019.2919404], https://hal.inria.fr/hal-01954261

[22] R. GRIBONVAL, M. NIKOLOVA.*On bayesian estimation and proximity operators*, in "Applied and Computational Harmonic Analysis", 2019, p. 1-25, https://arxiv.org/abs/1807.04021 , forthcoming [*DOI :* 10.1016/J.ACHA.2019.07.002], https://hal.inria.fr/hal-01835108

Invited Conferences

[23] R. GRIBONVAL, G. KUTYNIOK, M. NIELSEN, F. VOIGTLAENDER.*Approximation spaces of deep neural networks*, in "SMAI 2019 - 9ème Biennale des Mathématiques Appliquées et Industrielles", Guidel, France, May 2019, 1, https://hal.inria.fr/hal-02127179

[24] V. SCHELLEKENS, A. CHATALIC, F. HOUSSIAU, Y.-A. DE MONTJOYE, L. JACQUES, R. GRIBONVAL.*Differentially Private Compressive k-Means*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech, and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, p. 7933-7937 [*DOI :* 10.1109/ICASSP.2019.8682829], https://hal.inria.fr/hal-02060208

International Conferences with Proceedings

[25] A. CHATALIC, N. KERIVEN, R. GRIBONVAL.*Projections aléatoires pour l'apprentissage compressif*, in "GRETSI 2019 - XXVIIème Colloque francophone de traitement du signal et des images", Lille, France, August 2019, p. 1-4, https://hal.inria.fr/hal-02154803

[26] J. E. COHEN, N. GILLIS.*Nonnegative Low-rank Sparse Component Analysis*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, p. 8226-8230 [*DOI :* 10.1109/ICASSP.2019.8682188], https://hal.archives-ouvertes.fr/hal-02201471

[27] C. F. DANTAS, J. E. COHEN, R. GRIBONVAL.*Hyperspectral Image Denoising using Dictionary Learning*, in "WHISPERS 2019 - 10th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing", Amsterdam, Netherlands, September 2019, p. 1-5, https://hal.inria.fr/hal-02175630

[28] C. F. DANTAS, J. E. COHEN, R. GRIBONVAL.*Learning Tensor-structured Dictionaries with Application to Hyperspectral Image Denoising*, in "EUSIPCO 2019 - 27th European Signal Processing Conference", A Coruña, Spain, September 2019, p. 1-5, https://hal.inria.fr/hal-02126782

[29] D. DI CARLO, A. DELEFORGE, N. BERTIN.*Mirage: 2D Source Localization Using Microphone Pair Augmentation with Echoes*, in "ICASSP 2019 - IEEE International Conference on Acoustic, Speech Signal

Processing", Brighton, United Kingdom, IEEE, May 2019, p. 775-779, https://arxiv.org/abs/1906.08968 [*DOI :* 10.1109/ICASSP.2019.8683534], https://hal.archives-ouvertes.fr/hal-02160940

[30] C. ELVIRA, R. GRIBONVAL, C. HERZET, C. SOUSSEN.*Uniform k-step recovery with CMF dictionaries*, in "SPARS 2019 - Signal Processing with Adaptive Sparse Structured Representations", Toulouse, France, July 2019, p. 1-2, https://hal.inria.fr/hal-02157561

[31] C. ELVIRA, R. GRIBONVAL, C. SOUSSEN, C. HERZET.*OMP and continuous dictionaries: Is k-step recovery possible ?*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, p. 1-5 [*DOI :* 10.1109/ICASSP.2019.8683617], https://hal.archives-ouvertes.fr/hal-02049486

[32] V. GILLOT, F. BIMBOT.*Polytopic reconfiguration: a graph-based scheme for the multiscale transformation of music segments and its perceptual assessment*, in "SMC 2019 - 16th Sound & Music Computing Conference", Malaga, Spain, May 2019, p. 1-8, https://hal.archives-ouvertes.fr/hal-02132955

[33] S. GUPTA, R. GRIBONVAL, L. DAUDET, I. DOKMANIĆ.*Don't take it lightly: Phasing optical random projections with unknown operators*, in "NeurIPS 2019 - Thirty-third Conference on Neural Information Processing Systems", Vancouver, Canada, December 2019, p. 1-13, https://arxiv.org/abs/1907.01703 , https://hal.inria.fr/hal-02342280

[34] M. HAFSATI, N. EPAIN, R. GRIBONVAL, N. BERTIN.*Sound source separation in the higher order ambisonics domain*, in "DAFx 2019 - 22nd International Conference on Digital Audio Effects", Birmingham, United Kingdom, September 2019, p. 1-7, https://hal.inria.fr/hal-02161949

[35] V. SCHELLEKENS, A. CHATALIC, F. HOUSSIAU, Y.-A. DE MONTJOYE, L. JACQUES, R. GRIBONVAL.*Compressive k-Means with Differential Privacy*, in "SPARS 2019 - Signal Processing with Adaptive Sparse Structured Representations", Toulouse, France, July 2019, p. 1-2, https://hal.inria.fr/hal-02154820

National Conferences with Proceeding

[36] C. ELVIRA, R. GRIBONVAL, C. SOUSSEN, C. HERZET.*Identification de supports en k e´tapes avec OMP pour les dictionnaires continus*, in "GRETSI 2019 - XXVIIème Colloque francophone de traitement du signal et des images", Lille, France, August 2019, p. 1-4, https://hal.inria.fr/hal-02157571

Conferences without Proceedings

[37] A. ANG, J. E. COHEN, N. GILLIS.*Accelerating Approximate Nonnegative Canonical Polyadic Decomposition using Extrapolation*, in "GRETSI 2019 - XXVIIème Colloque francophone de traitement du signal et des images", Lille, France, August 2019, p. 1-4, https://hal.archives-ouvertes.fr/hal-02143969

[38] C. CURY, P. MAUREL, R. GRIBONVAL, C. BARILLOT.*Can we learn from coupling EEG-fMRI to enhance neuro-feedback in EEG only?*, in "OHBM 2019 - Annual Meeting Organization for Human Brain Mapping", Rome, Italy, June 2019, 1, https://www.hal.inserm.fr/inserm-02074623

[39] C. CURY, P. MAUREL, G. LIOI, R. GRIBONVAL, C. BARILLOT.*Learning bi-modal EEG-fMRI neurofeedback to improve neurofeedback in EEG only*, in "Real-Time Functional Imaging and Neurofeedback", Maastricht, Netherlands, December 2019, p. 1-2 [*DOI :* 10.1101/599589], https://www.hal.inserm.fr/inserm-02368720

[40] A. LORENTE MUR, M. OCHOA, J. E. COHEN, X. INTES, N. DUCROS.*Handling negative patterns for fast single-pixel lifetime imaging*, in "2019 - Molecular-Guided Surgery: Molecules, Devices, and Applications V", San Francisco, United States, SPIE, February 2019, vol. 10862, p. 1-10 [*DOI :* 10.1117/12.2511123], https://hal.archives-ouvertes.fr/hal-02017598

[41] P. STOCK, B. GRAHAM, R. GRIBONVAL, H. JÉGOU.*Equi-normalization of Neural Networks*, in "ICLR 2019 - Seventh International Conference on Learning Representations", New Orleans, United States, May 2019, p. 1-20, https://arxiv.org/abs/1902.10416 , https://hal.archives-ouvertes.fr/hal-02050408

[42] P. STOCK, A. JOULIN, R. GRIBONVAL, B. GRAHAM, H. JÉGOU.*And the Bit Goes Down: Revisiting the Quantization of Neural Networks*, in "ICLR 2020 - Eighth International Conference on Learning Representations", Addis-Abeba, Ethiopia, February 2020, p. 1-11, https://hal.archives-ouvertes.fr/hal-02434572

[43] T. VAYER, R. FLAMARY, R. TAVENARD, L. CHAPEL, N. COURTY.*Sliced Gromov-Wasserstein*, in "NeurIPS 2019 - Thirty-third Conference on Neural Information Processing Systems", Vancouver, Canada, December 2019, vol. 32, https://arxiv.org/abs/1905.10124 , https://hal.archives-ouvertes.fr/hal-02174309

Scientific Books (or Scientific Book chapters)

[44] D. K. HAMMOND, P. VANDERGHEYNST, R. GRIBONVAL.*The Spectral Graph Wavelet Transform: Fundamental Theory and Fast Computation*, in "Vertex-Frequency Analysis of Graph Signals", L. STANKOVIĆ, E. SEJDIĆ (editors), Signals and Communication Technology, Springer International Publishing, December 2019, p. 141-175 [*DOI :* 10.1007/978-3-030-03574-7_3], https://hal.inria.fr/hal-01943589

Research Reports

[45] B. CARAMIAUX, F. LOTTE, J. GEURTS, G. AMATO, M. BEHRMANN, F. BIMBOT, F. FALCHI, A. GARCIA, J. GIBERT, G. GRAVIER, H. HOLKEN, H. KOENITZ, S. LEFEBVRE, A. LIUTKUS, A. PERKIS, R. REDONDO, E. TURRIN, T. VIÉVILLE, E. VINCENT.*AI in the media and creative industries*, New European Media (NEM), April 2019, p. 1-35, https://arxiv.org/abs/1905.04175 , https://hal.inria.fr/hal-02125504

Software

[46] C. F. DANTAS, R. GRIBONVAL.*Stable Screening - Python code*, June 2019
[SWH-ID : swh:1:dir:3ccb3fde105b05aee192367fb5e07e192e3b6774], Software, https://hal.inria.fr/hal-02129219

[47] M. KOWALSKI, E. VINCENT, R. GRIBONVAL.*Underdetermined Reverberant Source Separation*, October 2019
[SWH-ID : swh:1:dir:ec4ae097465d9ea51589537ea94b2ea50e8d134d], Software, https://hal.archives-ouvertes.fr/hal-02309043

Other Publications

[48] A. ACKAOUY, N. COURTY, E. VALLEE, O. COMMOWICK, C. BARILLOT, F. GALASSI.*Unsupervised Domain Adaptation with Optimal Transport in multi-site segmentation of Multiple Sclerosis lesions from MRI data : Preprint*, October 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02317028

[49] B. B. DAMODARAN, R. FLAMARY, V. SEGUY, N. COURTY.*An Entropic Optimal Transport Loss for Learning Deep Neural Networks under Label Noise in Remote Sensing Images*, July 2019, https://arxiv.org/abs/1810.

01163 - Under Consideration at Computer Vision and Image Understanding [*DOI :* 10.01163], https://hal.archives-ouvertes.fr/hal-02174320

[50] C. F. DANTAS, J. E. COHEN, R. GRIBONVAL.*Tensor-structured Dictionaries for Hyperspectral Imaging*, July 2019, p. 1-2, SPARS 2019 - Signal Processing with Adaptive Sparse Structured Representations, https://hal.inria.fr/hal-02169405

[51] C. ELVIRA, R. GRIBONVAL, C. SOUSSEN, C. HERZET.*When does OMP achieve support recovery with continuous dictionaries?*, April 2019, working paper or preprint, https://hal.inria.fr/hal-02099464

[52] C. ELVIRA, C. HERZET.*Safe Squeezing for Antisparse Coding*, November 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02368134

[53] R. GRIBONVAL, G. KUTYNIOK, M. NIELSEN, F. VOIGTLAENDER.*Approximation spaces of deep neural networks*, June 2019, https://arxiv.org/abs/1905.01208 - working paper or preprint, https://hal.inria.fr/hal-02117139

[54] R. GRIBONVAL, M. NIKOLOVA.*A characterization of proximity operators*, November 2019, https://arxiv.org/abs/1807.04014 - working paper or preprint, https://hal.inria.fr/hal-01835101

[55] A. LORENTE MUR, J. E. COHEN, N. DUCROS.*Factorisation exacte pour la généralisation de motifs en imagerie monopixel*, March 2019, 1, JIONC - 14ème Journées d'Imagerie Optique Non Conventionelle, Poster, https://hal.archives-ouvertes.fr/hal-02170148

[56] T. VAYER, L. CHAPEL, R. FLAMARY, R. TAVENARD, N. COURTY.*Fused Gromov-Wasserstein distance for structured objects: theoretical foundations and mathematical properties*, July 2019, https://arxiv.org/abs/1811.02834 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02174316

References in notes

[57] 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^{o} 3, p. 922 - 932 [*DOI :* 10.1109/TASL.2011.2168211], http://hal.inria.fr/inria-00577079

[58] 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

[59] 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^{o} 19, 20 [*DOI :* 10.1109/TSP.2015.2447503], https://hal.archives-ouvertes.fr/hal-01084986

[60] A. BOURRIER.*Compressed sensing and dimensionality reduction for unsupervised learning*, Université Rennes 1, May 2014, https://tel.archives-ouvertes.fr/tel-01023030

[61] 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

[62] M. CHAFII, J. PALICOT, R. GRIBONVAL.*Dispositif de communication à modulation temps-fréquence adaptative*, July 2016, n^{o} Numéro de demande : 1656806 ; Numéro de soumission : 1000356937, https://hal.inria.fr/hal-01375661

[63] A. CHATALIC, R. GRIBONVAL, N. KERIVEN.*Large-Scale High-Dimensional Clustering with Fast Sketching*, in "ICASSP 2018 - IEEE International Conference on Acoustics, Speech and Signal Processing", Calgary, Canada, IEEE, April 2018, p. 4714-4718 [*DOI :* 10.1109/ICASSP.2018.8461328], https://hal.inria.fr/hal-01701121

[64] I. DOKMANIĆ, R. GRIBONVAL.*Beyond Moore-Penrose Part I: Generalized Inverses that Minimize Matrix Norms*, July 2017, working paper or preprint, https://hal.inria.fr/hal-01547183

[65] I. DOKMANIĆ, R. PARHIZKAR, A. WALTHER, Y. M. LU, M. VETTERLI.*Acoustic echoes reveal room shape*, in "Proceedings of the National Academy of Sciences", 2013, vol. 110, n^{o} 30, p. 12186–12191, http://dx.doi.org/10.1073/pnas.1221464110

[66] C. ELVIRA, R. GRIBONVAL, C. HERZET, C. SOUSSEN.*A case of exact recovery with OMP using continuous dictionaries*, in "CS 2018 - 9th International Conference on Curves and Surfaces", Arcachon, France, June 2018, https://hal.inria.fr/hal-01937532

[67] C. FRAGA DANTAS, J. E. COHEN, R. GRIBONVAL.*Learning fast dictionaries for sparse representations using low-rank tensor decompositions*, in "LVA/ICA 2018 - 14th International Conference on Latent Variable Analysis and Signal Separation", Guildford, United Kingdom, LNCS, Springer, July 2018, vol. 10891, p. 456-466 [*DOI :* 10.1007/978-3-319-93764-9_42], https://hal.inria.fr/hal-01709343

[68] 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

[69] C. GAULTIER, N. BERTIN, S. KITIĆ, R. GRIBONVAL.*A modeling and algorithmic framework for (non)social (co)sparse audio restoration*, November 2017, working paper or preprint, https://hal.inria.fr/hal-01649261

[70] M. A. GERZON.*Periphony: With-Height Sound Reproduction*, in "J. Audio Eng. Soc", 1973, vol. 21, n^{o} 1, p. 2–10, http://www.aes.org/e-lib/browse.cfm?elib=2012

[71] R. GRIBONVAL, G. BLANCHARD, N. KERIVEN, Y. TRAONMILIN.*Compressive Statistical Learning with Random Feature Moments*, December 2017, 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

[72] R. GRIBONVAL.*Should penalized least squares regression be interpreted as Maximum A Posteriori estimation?*, in "IEEE Transactions on Signal Processing", May 2011, vol. 59, n^{o} 5, p. 2405-2410 [*DOI :* 10.1109/TSP.2011.2107908], http://hal.inria.fr/inria-00486840

[73] H. JAIN.*Learning compact representations for large scale image search*, Université Rennes 1, June 2018, https://tel.archives-ouvertes.fr/tel-01889405

[74] N. KERIVEN.*Sketching for large-scale learning of mixture models*, Université Rennes 1, October 2017, https://tel.archives-ouvertes.fr/tel-01620815

[75] 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

[76] L. Le Magoarou, R. Gribonval, N. Tremblay.*Approximate fast graph Fourier transforms via multi-layer sparse approximations*, in "IEEE transactions on Signal and Information Processing over Networks", June 2018, vol. 4, n^{o} 2, p. 407–420, https://arxiv.org/abs/1612.04542 [*DOI :* 10.1109/TSIPN.2017.2710619], https://hal.inria.fr/hal-01416110

[77] R. Lebarbenchon, E. Camberlein, D. Di Carlo, C. Gaultier, A. Deleforge, N. Bertin.*Evaluation of an Open-Source Implementation of the SRP-PHAT Algorithm within the 2018 Locata Challenge*, in "LOCATA Challenge Workshop, a satellite event of IWAENC 2018", Tokyo, Japan, September 2018, https://hal.archives-ouvertes.fr/hal-02187964

[78] N. Libermann, F. Bimbot, E. Vincent.*Exploration de dépendances structurelles mélodiques par réseaux de neurones récurrents*, in "JIM 2018 - Journées d'Informatique Musicale", Amiens, France, May 2018, p. 81-86, https://hal.archives-ouvertes.fr/hal-01791381

[79] 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

[80] 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

[81] 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

[82] 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

[83] H. Peic Tukuljac, A. Deleforge, R. Gribonval.*MULAN: A Blind and Off-Grid Method for Multichannel Echo Retrieval*, in "NIPS 2018 - Thirty-second Conference on Neural Information Processing Systems", Montréal, Canada, December 2018, p. 1-11, https://arxiv.org/abs/1810.13338 , https://hal.inria.fr/hal-01906385

[84] 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

[85] R. Scheibler, D. Di Carlo, A. Deleforge, I. Dokmanić.*Separake: Source Separation with a Little Help From Echoes*, in "ICASSP 2018 - IEEE International Conference on Acoustics, Speech and Signal Processing", Calgary, Canada, April 2018, https://hal.inria.fr/hal-01909531

Project-Team RAINBOW

Sensor-based Robotics and Human Interaction

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
Robotics and Smart environments

Table of contents

Project-Team RAINBOW

Creation of the Team: 2018 January 01, updated into Project-Team: 2018 June 01

Keywords:

Computer Science and Digital Science:

A5.1.3. - Haptic interfaces
A5.1.7. - Multimodal interfaces
A5.4.4. - 3D and spatio-temporal reconstruction
A5.4.6. - Object localization
A5.4.7. - Visual servoing
A5.5.4. - Animation
A5.6. - Virtual reality, augmented reality
A5.6.1. - Virtual reality
A5.6.2. - Augmented reality
A5.6.3. - Avatar simulation and embodiment
A5.6.4. - Multisensory feedback and interfaces
A5.9.2. - Estimation, modeling
A5.10.2. - Perception
A5.10.3. - Planning
A5.10.4. - Robot control
A5.10.5. - Robot interaction (with the environment, humans, other robots)
A5.10.6. - Swarm robotics
A5.10.7. - Learning
A6.4.1. - Deterministic control
A6.4.3. - Observability and Controlability
A6.4.4. - Stability and Stabilization
A6.4.5. - Control of distributed parameter systems
A6.4.6. - Optimal control
A9.5. - Robotics
A9.7. - AI algorithmics
A9.9. - Distributed AI, Multi-agent

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
B8.1.2. - Sensor networks for smart buildings
B8.4. - Security and personal assistance

1. Team, Visitors, External Collaborators

Research Scientists

Paolo Robuffo Giordano [Team leader, CNRS, Senior Researcher, HDR]
François Chaumette [Inria, Senior Researcher, HDR]
Alexandre Krupa [Inria, Researcher, HDR]
Claudio Pacchierotti [CNRS, Researcher]
Julien Pettré [Inria, Senior Researcher, HDR]

Faculty Members

Marie Babel [INSA Rennes, Associate Professor, HDR]
Vincent Drevelle [Univ de Rennes I, Associate Professor]
Eric Marchand [Univ de Rennes I, Professor, HDR]
Quentin Delamare [ENS Rennes, Professor, AGPR, from Sep 2019]

Technical Staff

Fabien Spindler [Inria, Engineer]
Firas Abi Farraj [CNRS, until Jun 2019]
Marco Aggravi [CNRS]
Dieudonne Atrevi [Inria, from Nov 2019]
Arianna Bottinelli [Inria, from Feb 2019 until May 2019]
Julien Bruneau [Inria]
Nicolas Cazy [Inria, until Feb 2019]
Marco Cognetti [CNRS, until Nov 2019]
Cedric de Almeida Braga [Inria, from Oct 2019]
Louise Devigne [INSA Rennes]
Solenne Fortun [INSA Rennes]
Ceilidh Hoffmann [Inria, until Mar 2019]
Joudy Nader [CNRS, from Oct 2019]
Noura Neji [Inria, from Sep 2019]
François Pasteau [INSA Rennes]
Ramana Sundararaman [Inria, from Oct 2019]
Wouter Van Toll [Inria, from Jan 2019]

PhD Students

Javad Amirian [Inria]
Benoît Antoniotti [Inria]
Florian Berton [Inria]
Pascal Brault [Inria, from Sep 2019]
Adèle Colas [Inria, from Nov 2019]
Quentin Delamare [Univ de Rennes I, until Aug 2019]
Mathieu Gonzalez [IRT B<>Com, from Oct 2019]
Fabien Grzeskowiak [Inria]
Hadrien Gurnel [IRT B<>com, until Oct 2019]
Alberto Jovane [Inria, from Sep 2019]
Axel Lopez Gandia [Univ de Rennes I]
Alexander Oliva [Inria]
Rahaf Rahal [Univ de Rennes I]
Agniva Sengupta [Inria]
Guillaume Vailland [INSA Rennes]

Post-Doctoral Fellow

Thomas Howard [CNRS]

Visiting Scientist

Ekrem Misimi [Sintef, Norway, Mar 2019 and July 2019]

Administrative Assistant

Hélène de La Ruée [Inria]

2. Overall Objectives

2.1. Overall Objectives

The long-term vision of the Rainbow team is to develop the next generation of sensor-based robots able to navigate and/or interact in complex unstructured environments *together* with human users. Clearly, the word "together" can have very different meanings depending on the particular context: for example, it can refer to mere co-existence (robots and humans share some space while performing independent tasks), human-awareness (the robots need to be aware of the human state and intentions for properly adjusting their actions), or actual cooperation (robots and humans perform some shared task and need to coordinate their actions).

One could perhaps argue that these two goals are somehow in conflict since higher robot autonomy should imply lower (or absence of) human intervention. However, we believe that our general research direction is well motivated since: (i) despite the many advancements in robot autonomy, complex and high-level cognitive-based decisions are still out of reach. In most applications involving tasks in unstructured environments, uncertainty, and interaction with the physical word, human assistance is still necessary, and will most probably be for the next decades. On the other hand, robots are extremely capable at autonomously executing specific and repetitive tasks, with great speed and precision, and at operating in dangerous/remote environments, while humans possess unmatched cognitive capabilities and world awareness which allow them to take complex and quick decisions; (ii) the cooperation between humans and robots is often an implicit constraint of the robotic task itself. Consider for instance the case of assistive robots supporting injured patients during their physical recovery, or human augmentation devices. It is then important to study proper ways of implementing this cooperation; (iii) finally, safety regulations can require the presence at all times of a person in charge of supervising and, if necessary, take direct control of the robotic workers. For example, this is a common requirement in all applications involving tasks in public spaces, like autonomous vehicles in crowded spaces, or even UAVs when flying in civil airspace such as over urban or populated areas.

Within this general picture, the Rainbow activities will be particularly focused on the case of **(shared) cooperation between robots and humans** by pursuing the following vision: on the one hand, empower robots with a *large degree of autonomy* for allowing them to effectively operate in non-trivial environments (e.g., outside completely defined factory settings). On the other hand, include *human users* in the loop for having them in (partial and bilateral) control of some aspects of the overall robot behavior. We plan to address these challenges from the **methodological**, **algorithmic** and **application-oriented** perspectives. The main *research axes* along which the Rainbow activities will be articulated are: three supporting axes (**Optimal and Uncertainty-Aware Sensing**; **Advanced Sensor-based Control**; **Haptics for Robotics Applications**) that are meant to develop methods, algorithms and technologies for realizing the central theme of **Shared Control of Complex Robotic Systems**.

3. Research Program

3.1. Main Vision

The vision of Rainbow (and foreseen applications) calls for several general scientific challenges: (i) high-level of autonomy for complex robots in complex (unstructured) environments, (ii) forward interfaces for letting an operator giving high-level commands to the robot, (iii) backward interfaces for informing the operator about the robot 'status', (iv) user studies for assessing the best interfacing, which will clearly depend on the particular task/situation. Within Rainbow we plan to tackle these challenges at different levels of depth:

- the **methodological and algorithmic side** of the sought human-robot interaction will be the **main focus** of Rainbow. Here, we will be interest in advancing the state-of-the-art in sensor-based online planning, control and manipulation for mobile/fixed robots. For instance, while classically most control approaches (especially those sensor-based) have been essentially *reactive*, we believe that less myopic strategies based on online/reactive trajectory optimization will be needed for the future

Rainbow activities. The core ideas of Model-Predictive Control approaches (also known as Receding Horizon) or, in general, numerical optimal control methods will play a role in the Rainbow activities, for allowing the robots to reason/plan over some future time window and better cope with constraints. We will also consider extending classical sensor-based motion control/manipulation techniques to more realistic scenarios, such as deformable/flexible objects ("**Advanced Sensor-based Control**" axis). Finally, it will also be important to spend research efforts into the field of *Optimal Sensing*, in the sense of generating (again) trajectories that can optimize the state estimation problem in presence of scarce sensory inputs and/or non-negligible measurement and process noises, especially true for the case of mobile robots ("**Optimal and Uncertainty-Aware Sensing**" axis). We also aim at addressing the case of coordination between a single human user and multiple robots where, clearly, as explained the autonomy part plays even a more crucial role (no human can control multiple robots at once, thus a high degree of autonomy will be required by the robot group for executing the human commands);

- the **interfacing side** will also be a focus of the Rainbow activities. As explained above, we will be interested in both the *forward* (human $\rightarrow$ robot) and *backward* (robot $\rightarrow$ human) interfaces. The forward interface will be mainly addressed from the *algorithmic* point of view, i.e., how to map the few degrees of freedom available to a human operator (usually in the order of 3–4) into complex commands for the controlled robot(s). This mapping will typically be mediated by an "AutoPilot" onboard the robot(s) for autonomously assessing if the commands are feasible and, if not, how to least modify them ("**Advanced Sensor-based Control**" axis).
  The backward interface will, instead, mainly consist of a visual/haptic feedback for the operator. Here, we aim at exploiting our expertise in using force cues for informing an operator about the status of the remote robot(s). However, the sole use of classical *grounded* force feedback devices (e.g., the typical force-feedback joysticks) will not be enough due to the different kinds of information that will have to be provided to the operator. In this context, the recent interest in the use of *wearable* haptic interfaces is very interesting and will be investigated in depth (these include, e.g., devices able to provide vibro-tactile information to the fingertips, wrist, or other parts of the body). The main challenges in these activities will be the mechanical conception (and construction) of suitable wearable interfaces for the tasks at hand, and in the generation of force cues for the operator: the force cues will be a (complex) function of the robot state, therefore motivating research in algorithms for mapping the robot state into a few variables (the force cues) ("**Haptics for Robotics Applications**" axis);
- the **evaluation side** that will assess the proposed interfaces with some user studies, or acceptability studies by human subjects. Although this activity **will not** be a main focus of Rainbow (complex user studies are beyond the scope of our core expertise), we will nevertheless devote some efforts into having some reasonable level of user evaluations by applying standard statistical analysis based on psychophysical procedures (e.g., randomized tests and Anova statistical analysis). This will be particularly true for the activities involving the use of smart wheelchairs, which are intended to be used by human users *and* operate inside human crowds. Therefore, we will be interested in gaining some level of understanding of how semi-autonomous robots (a wheelchair in this example) can predict the human intention, and how humans can react to a semi-autonomous mobile robot.

Figure 1 depicts in an illustrative way the *prototypical* activities foreseen in Rainbow. On the righthand side, complex robots (dual manipulators, humanoid, single/multiple mobile robots) need to perform some task with high degree of autonomy. On the lefthand side, a human operator gives some high-level commands and receives a visual/haptic feedback aimed at informing her/him at best of the robot status. Again, the main challenges that Rainbow will tackle to address these issues are (in order of relevance): (i) methods and algorithms, mostly based on first-principle modeling and, when possible, on numerical methods for online/reactive trajectory generation, for enabling the robots with high autonomy; (ii) design and implementation of visual/haptic cues for interfacing the human operator with the robots, with a special attention to novel combinations of grounded/ungrounded (wearable) haptic devices; (iii) user and acceptability studies.

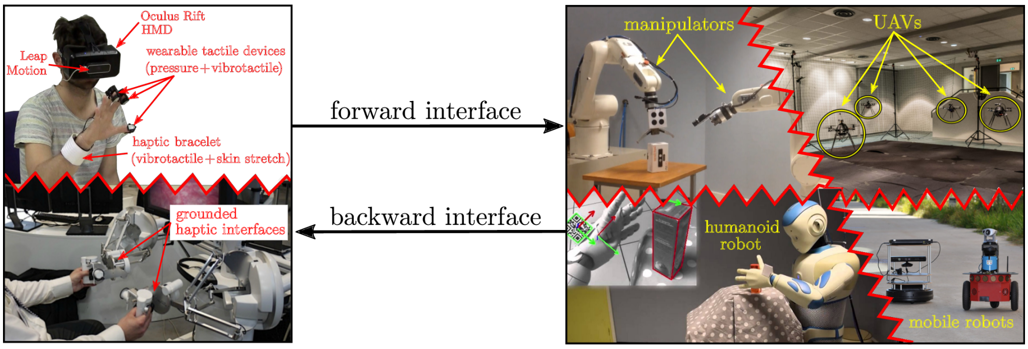


Figure 1. An illustration of the prototypical activities foreseen in Rainbow in which a human operator is in partial (and high-level) control of single/multiple complex robots performing semi-autonomous tasks

3.2. Main Components

Hereafter, a summary description of the four axes of research in Rainbow.

3.2.1. Optimal and Uncertainty-Aware Sensing

Future robots will need to have a large degree of autonomy for, e.g., interpreting the sensory data for accurate estimation of the robot and world state (which can possibly include the human users), and for devising motion plans able to take into account many constraints (actuation, sensor limitations, environment), including also the state estimation accuracy (i.e., how well the robot/environment state can be reconstructed from the sensed data). In this context, we will be particularly interested in (i) devising trajectory optimization strategies able to maximize some norm of the information gain gathered along the trajectory (and with the available sensors). This can be seen as an instance of Active Sensing, with the main focus on *online/reactive* trajectory optimization strategies able to take into account several requirements/constraints (sensing/actuation limitations, noise characteristics). We will also be interested in the coupling between optimal sensing and concurrent execution of additional tasks (e.g., navigation, manipulation). (ii) Formal methods for guaranteeing the accuracy of localization/state estimation in mobile robotics, mainly exploiting tools from interval analysis. The interest in these methods is their ability to provide possibly conservative but guaranteed accuracy bounds on the best accuracy one can obtain with the given robot/sensor pair, and can thus be used for planning purposes of for system design (choice of the best sensor suite for a given robot/task). (iii) Localization/tracking of objects with poor/unknown or deformable shape, which will be of paramount importance for allowing robots to estimate the state of "complex objects" (e.g., human tissues in medical robotics, elastic materials in manipulation) for controlling its pose/interaction with the objects of interest.

3.2.2. Advanced Sensor-based Control

One of the main competences of the previous Lagadic team has been, generally speaking, the topic of *sensor-based control*, i.e., how to exploit (typically onboard) sensors for controlling the motion of fixed/ground robots. The main emphasis has been in devising ways to directly couple the robot motion with the sensor outputs in order to invert this mapping for driving the robots towards a configuration specified as a desired sensor reading (thus, directly in sensor space). This general idea has been applied to very different contexts: mainly standard vision (from which the Visual Servoing keyword), but also audio, ultrasound imaging, and RGB-D.

Use of sensors for controlling the robot motion will also clearly be a central topic of the Rainbow team too, since the use of (especially onboard) sensing is a main characteristics of any future robotics application (which should typically operate in unstructured environments, and thus mainly rely on its own ability to sense the world). We then naturally aim at making the best out of the previous Lagadic experience in sensor-based control for proposing new advanced ways of exploiting sensed data for, roughly speaking, controlling the motion of a robot. In this respect, we plan to work on the following topics: (i) "direct/dense methods" which try to directly exploit the raw sensory data in computing the control law for positioning/navigation tasks. The advantages of these methods is the need for little data pre-processing which can minimize feature extraction errors and, in general, improve the overall robustness/accuracy (since all the available data is used by the motion controller); (ii) sensor-based interaction with objects of unknown/deformable shapes, for gaining the ability to manipulate, e.g., flexible objects from the acquired sensed data (e.g., controlling online a needle being inserted in a flexible tissue); (iii) sensor-based model predictive control, by developing *online/reactive* trajectory optimization methods able to plan feasible trajectories for robots subjects to sensing/actuation constraints with the possibility of (onboard) sensing for continuously replanning (over some future time horizon) the optimal trajectory. These methods will play an important role when dealing with complex robots affected by complex sensing/actuation constraints, for which pure reactive strategies (as in most of the previous Lagadic works) are not effective. Furthermore, the coupling with the aforementioned optimal sensing will also be considered; (iv) multi-robot decentralised estimation and control, with the aim of devising again sensor-based strategies for groups of multiple robots needing to maintain a formation or perform navigation/manipulation tasks. Here, the challenges come from the need of devising "simple" decentralized and scalable control strategies under the presence of complex sensing constraints (e.g., when using onboard cameras, limited fov, occlusions). Also, the need of locally estimating global quantities (e.g., common frame of reference, global property of the formation such as connectivity or rigidity) will also be a line of active research.

3.2.3. Haptics for Robotics Applications

In the envisaged *shared* cooperation between human users and robots, the typical sensory channel (besides vision) exploited to inform the human users is most often the force/kinesthetic one (in general, the sense of touch and of applied forces to the human hand or limbs). Therefore, a part of our activities will be devoted to study and advance the use of *haptic* cueing algorithms and interfaces for providing a feedback to the users during the execution of some shared task. We will consider: (i) multi-modal haptic cueing for general teleoperation applications, by studying how to convey information through the kinesthetic and cutaneous channels. Indeed, most haptic-enabled applications typically only involve kinesthetic cues, e.g., the forces/torques that can be felt by grasping a force-feedback joystick/device. These cues are very informative about, e.g., preferred/forbidden motion directions, but are also inherently limited in their resolution since the kinesthetic channel can easily become overloaded (when too much information is compressed in a single cue). In recent years, the arise of novel cutaneous devices able to, e.g., provide vibro-tactile feedback on the fingertips or skin, has proven to be a viable solution to *complement* the classical kinesthetic channel. We will then study how to combine these two sensory modalities for different prototypical application scenarios, e.g., 6-dof teleoperation of manipulator arms, virtual fixtures approaches, and remote manipulation of (possibly deformable) objects; (ii) in the particular context of medical robotics, we plan to address the problem of providing haptic cues for typical medical robotics tasks, such as semi-autonomous needle insertion and robot surgery by exploring the use of kinesthetic feedback for rendering the mechanical properties of the tissues, and vibrotactile feedback for providing with guiding information about pre-planned paths (with the aim of increasing the usability/acceptability of this technology in the medical domain); (iii) finally, in the context of multi-robot control we would like to explore how to use the haptic channel for providing information about the status of *multiple* robots executing a navigation or manipulation task. In this case, the problem is (even more) how to map (or compress) information about many robots into a few haptic cues. We plan to use specialized devices, such as actuated exoskeleton gloves able to provide cues to each fingertip of a human hand, or to resort to "compression" methods inspired by the *hand postural synergies* for providing coordinated cues representative of a few (but complex) motions of the multi-robot group, e.g., coordinated motions (translations/expansions/rotations) or collective grasping/transporting.

3.2.4. Shared Control of Complex Robotics Systems

This final and main research axis will exploit the **methods, algorithms and technologies** developed in the previous axes for realizing applications involving complex semi-autonomous robots operating in complex environments together with human users. The *leitmotiv* is to realize advanced *shared control* paradigms, which essentially aim at blending robot autonomy and user's intervention in an optimal way for exploiting the best of both worlds (robot accuracy/sensing/mobility/strength and human's cognitive capabilities). A common theme will be the issue of where to "draw the line" between robot autonomy and human intervention: obviously, there is no general answer, and any design choice will depend on the particular task at hand and/or on the technological/algorithmic possibilities of the robotic system under consideration.

A *prototypical* envisaged application, exploiting and combining the previous three research axes, is as follows: a complex robot (e.g., a two-arm system, a humanoid robot, a multi-UAV group) needs to operate in an environment exploiting its onboard sensors (in general, vision as the main exteroceptive one) and deal with many constraints (limited actuation, limited sensing, complex kinematics/dynamics, obstacle avoidance, interaction with difficult-to-model entities such as surrounding people, and so on). The robot must then possess a quite large autonomy for interpreting and exploiting the sensed data in order to estimate its own state and the environment one ("**Optimal and Uncertainty-Aware Sensing**" axis), and for planning its motion in order to fulfil the task (e.g., navigation, manipulation) by coping with all the robot/environment constraints. Therefore, advanced control methods able to exploit the sensory data at its most, and able to cope *online* with constraints in an optimal way (by, e.g., continuously replanning and predicting over a future time horizon) will be needed ("**Advanced Sensor-based Control**" axis), with a possible (and interesting) coupling with the sensing part for optimizing, at the same time, the state estimation process. Finally, a human operator will typically be in charge of providing high-level commands (e.g., where to go, what to look at, what to grasp and where) that will then be autonomously executed by the robot, with possible local modifications because of the various (local) constraints. At the same time, the operator will also receive *online* visual-force cues informative of, in general, how well her/his commands are executed and if the robot would prefer or suggest other plans (because of the local constraints that are not of the operator's concern). This information will have to be visually and haptically rendered with an optimal combination of cues that will depend on the particular application ("**Haptics for Robotics Applications**" axis).

4. Highlights of the Year

4.1. Highlights of the Year

- J. Pettré is the unit PI of the new H2020 ICT project "PRESENT" started on Sep 2019

4.1.1. Awards

- P. Robuffo Giordano received the Prix Michel Monpetit – Inria from the Académie des sciences
- B. Penin (former PhD student), P. Robuffo Giordano and F. Chaumette received at ICRA 2019 the IEEE RA-L 2018 Best Paper Award for the paper "Vision-Based Reactive Planning for Aggressive Target Tracking while Avoiding Collisions and Occlusions"
- M. Babel received the Innovation Award from the Société Française de Médecine physique et de Réadaptation (SOFMER) for the power wheelchair simulator in virtual reality described in Sect. 6.4.6

5. New Software and Platforms

5.1. 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
- Partner: INSA Rennes
- Contact: Marie Babel

5.2. UsTk

Ultrasound toolkit for medical robotics applications guided from ultrasound images

KEYWORDS: Echographic imagery - Image reconstruction - Medical robotics - Visual tracking - Visual servoing (VS) - Needle insertion

FUNCTIONAL DESCRIPTION: UsTK, standing for Ultrasound Toolkit, is a cross-platform extension of ViSP software dedicated to 2D and 3D 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 acquiring 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 of ultrasound images, generate elastography images, track a flexible needle in sequences of 2D and 3D ultrasound images and track a target image template in sequences of 2D ultrasound images. All these modules were implemented on several robotic demonstrators to control the motion of an ultrasound probe or a flexible needle by ultrasound visual servoing.

- Participants: Alexandre Krupa and Fabien Spindler
- Partners: Inria - Université de Rennes 1
- Contact: Alexandre Krupa
- URL: https://ustk.inria.fr

5.3. ViSP

Visual servoing platform

KEYWORDS: Augmented reality - Computer vision - Robotics - Visual servoing (VS) - Visual tracking

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 Rainbow research. ViSP was designed to be independent from the hardware, to be simple to use, expandable and cross-platform. ViSP allows designing vision-based tasks for eye-in-hand and eye-to-hand systems from 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 300 fully documented classes organized in 17 different modules, with more than 408 examples and 88 tutorials are proposed to the user. ViSP is released under a dual licensing model. It is open-source with a GNU GPLv2 or GPLv3 license. A professional edition license that replaces GNU GPL is also available.

- Participants: Éric Marchand, Fabien Spindler and François Chaumette
- Partners: Inria - Université de Rennes 1
- Contact: Fabien Spindler
- URL: http://visp.inria.fr

5.4. Platforms

5.4.1. Robot Vision Platform

Participants: François Chaumette, Alexandre Krupa, Eric Marchand, Fabien Spindler [contact].

We exploit two industrial robotic systems built by Afma Robots in the nineties to validate our research 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). These robots are equipped with monocular RGB cameras. The Gantry robot also allows mounting grippers on its end-effector. Attached to this platform, we can also find a collection of various RGB and RGB-D cameras used to validate vision-based real-time tracking algorithms (see Sections 6.1.1 and 6.1.2). Note that four papers [32], [14], [52], [53] published by Rainbow in 2019 include results validated on this platform.

Figure 2. Rainbow robotics platform for vision-based manipulation

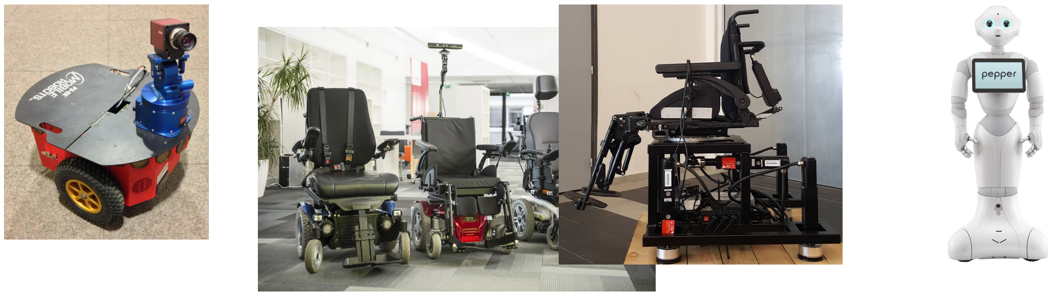
5.4.2. Mobile Robots

Participants: Marie Babel, Solenne Fortun, François Pasteau, Julien Pettré, Quentin Delamare, Fabien Spindler [contact].

For fast prototyping of algorithms in perception, control and autonomous navigation, the team uses a Pioneer 3DX from Adept (see Fig. 3.a). This platform is equipped with various sensors needed for autonomous navigation and sensor-based control.

Moreover, to validate our research in personally assisted living topic (see Section 6.4.4), we have three electric wheelchairs, one from Permobil, one from Sunrise and the last from YouQ (see Fig. 3.b). 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, ultrasound and time of flight sensors to perform the required servoing for assisting handicapped people. This year we also bought a wheelchair haptic simulator to develop new human interaction strategies in a virtual reality environment (see Fig. 3(c)).

Pepper, a human-shaped robot designed by SoftBank Robotics to be a genuine day-to-day companion (see Fig. 3.d) is also part of this platform. 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 robot-human interactions during locomotion (see Section 6.2.6).



(a) (b) (c) (d)

Figure 3. Mobile Robot Platform. a) Pioneer P3-DX robot, b) wheelchairs from Permobil, Sunrise and YouQ, c) Wheelchair haptic simulator, d) Pepper human-shaped robot

Note that 2 papers and 1 PhD Thesis exploiting the mobile robots were published this year [42], [54], [1].

5.4.3. Medical Robotic Platform

Participants: Alexandre Krupa, Fabien Spindler [contact].

This platform is composed of two 6 DoF Adept Viper arms (see Figs. 4.a–b). Ultrasound probes connected either to a SonoSite 180 Plus or an Ultrasonix SonixTouch 2D and 3D imaging system can be mounted on a force torque sensor attached to each robot end-effector. The haptic Virtuose 6D or Omega 6 device (see Fig. 7.a) can also be used within this platform.

This year this platform was extended with a new ATI Nano43 force/torque sensor dedicated to needle insertion applications.

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 Sect. 6.4.3 and Sect. 6.4.2).

This platform was used to obtain experimental results presented in 4 new papers [6], [23], [38], [51].

(a) (b)

Figure 4. Rainbow medical robotic platforms. a) 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.

5.4.4. Advanced Manipulation Platform

Participants: François Chaumette, Claudio Pacchierotti, Paolo Robuffo Giordano, Fabien Spindler [contact].

This new platform is composed by 2 Panda lightweight arms from Franka Emika equipped with torque sensors in all seven axes. An electric gripper, a camera or a soft hand from qbrobotics can be mounted on the robot end-effector (see Fig. 5.a) to validate our researches in coupling force and vision for controlling robot manipulators (see Section 6.2.12) and in shared control for remote manipulation (see Section 6.4.1). Other haptic devices (see Section 5.4.6) can also be coupled to this platform.

This year this platform was extended with a new Reflex TakkTile 2 gripper from RightHand Labs (see Fig. 5.b). A new force/torque sensor from Alberobotics that can be mounted on the robot end-effector to get more precision during torque control was also bought.

Two new papers published this year include experimental results obtained with this platform [50], [66].

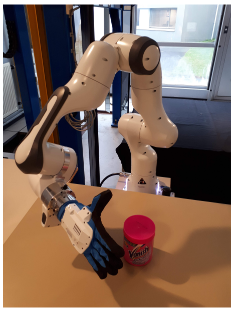


Figure 5. Rainbow advanced manipulation platform. a) One of the two Panda lightweight arms from Franka Emika, with mounted the Pisa SoftHand, b) the Reflex TakkTile 2 gripper that could be mounted on the Panda robot end-effector.

5.4.5. Unmanned Aerial Vehicles (UAVs)

Participants: Joudy Nader, Paolo Robuffo Giordano, Claudio Pacchierotti, Fabien Spindler [contact].

Rainbow is involved in several activities involving perception and control for single and multiple quadrotor UAVs. To this end, we purchased four quadrotors from Mikrokopter Gmbh, Germany (see Fig. 6.a), and one quadrotor from 3DRobotics, USA (see Fig. 6.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 a NVIDIA Jetson TX2 board running Linux Ubuntu and the TeleKyb-3 software based on genom3 framework developed at LAAS in Toulouse (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. 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 2 papers [20], [25] and 2 PhD Theses [3] contain simulative and experimental results obtained with this platform [1].

5.4.6. Haptics and Shared Control Platform

Participants: Claudio Pacchierotti, Paolo Robuffo Giordano, Fabien Spindler [contact].

Various haptic devices are used to validate our research in shared control. We have a Virtuose 6D device from Haption (see Fig. 7.a). This device is used as master device in many of our shared control activities (see

(a) (b) (c)

Figure 6. Unmanned Aerial Vehicles Platform. a) Quadrotor XL1 from Mikrokopter, b) Quadrotor Iris from 3DRobotics, c) Formation control with 3 XL1 from Mikrokopter.

Sections 6.4.2 and 6.4.3). It could also be coupled to the Haption haptic glove in loan from the University of Birmingham. An Omega 6 (see Fig. 7.b) from Force Dimension and devices in loan from Ultrahaptics complete this platform that could be coupled to the other robotic platforms.

This platform was used to obtain experimental results presented in 9 papers [6], [50], [43], [44], [38], [45], [11], [51], [5] published this year.

(a) (b)

Figure 7. Haptics and Shared Control Platform. a) Virtuose 6D and b) Omega 6 haptic devices

6. New Results

6.1. Optimal and Uncertainty-Aware Sensing

6.1.1. Tracking of Rigid Objects of Complex Shapes with a RDB-D Camera

Participants: Agniva Sengupta, Alexandre Krupa, Eric Marchand.

In the context of the iProcess project (see Section 8.3.8), we developed a method for accurately tracking the pose of rigid objects of complex shapes using a RGB-D camera [52]. This method only needs a coarse 3D geometric model of the object of interest represented as a 3D mesh. The tracking of the object is based on a joint minimization of geometric and photometric criteria and more particularly on a combination of point-to-plane distance minimization and photometric error minimization. The concept of successive "keyframes" was also used in this approach for minimizing possible drift of the tracking. The proposed approach was validated on both simulated and real data and the results experimentally demonstrated a better tracking accuracy than existing state-of-the-art 6-DoF object tracking methods, especially when dealing with low-textured objects, multiple coplanar faces, occlusions and partial specularities of the scene.

6.1.2. Deformable Object 3D Tracking based on Depth Information and Coarse Physical Model

Participants: Agniva Sengupta, Alexandre Krupa, Eric Marchand.

This research activity was also carried out in the context of the iProcess project (see Section 8.3.8) and will continue with the recent starting GentleMAN project (see Section 8.3.9). It focusses on the elaboration of approaches able to accurately track in real-time the deformation of soft objects using a RGB-D camera. The state-of-the-art approaches are currently relying on the use of Finite Element Model (FEM) to simulate the physics (mechanical behavior) of the deformable object. However, they suffer from the drawback of being excessively dependent on the accurate knowledge of the physical properties of the object being tracked (Young Modulus, Poisson's ratio, etc). This year, we proposed a first method that only required a coarse physical model of the object based on FEM whose parameters do not need to be precise [53]. The method consists in applying a set of virtual forces on the surface mesh of our coarse FEM model in such a way that it deforms to fit the current shape of the object. A point-to-plane distance error between the point cloud provided by the depth camera and the model mesh is iteratively minimized with respect to these virtual forces. The point of application of force is determined by an analysis of the error obtained from rigid tracking, which is done in parallel with the non-rigid tracking. The approach has been validated on simulated objects with ground-truth, as well on real objects of unknown physical properties and experimentally demonstrated that accurate tracking of deformable objects can be achieved without the need of a precise physical model.

6.1.3. Trajectory Generation for Optimal State Estimation

Participants: Marco Cognetti, Paolo Robuffo Giordano.

This activity addresses the general problem of *active sensing* where the goal is to analyze and synthesize optimal trajectories for a robotic system that can maximize the amount of information gathered by the (few) noisy outputs (i.e., sensor readings) while at the same time reducing the negative effects of the process/actuation noise. Over the last years we have developed a general framework for solving *online* the active sensing problem by continuously replanning an optimal trajectory that maximizes a suitable norm of the Constructibility Gramian (CG), while also coping with a number of constraints including limited energy and feasibility. The results obtained so far have been generalized and summarized in [27], where the online trajectory replanning for CG maximization has been applied to two relevant case studies (unicycle and quadrotor) and validated via a large statistical campaign. We are actually working towards the extension of this machinery to the case of realization of a robot task (e.g., reaching and grasping for a mobile manipulator), and to the mutual localization problem for a multi-robot group.

6.1.4. Robotic manipulators in Physical Interaction with the Environment

Participant: Claudio Pacchierotti.

As robotic systems become more flexible and intelligent, they must be able to move into environments with a high degree of uncertainty or clutter, such as our homes, workplaces, and the outdoors. In these unstructured scenarios, it is possible that the body of the robot collides with its surroundings. As such, it would be desirable to characterise these contacts in terms of their location and interaction forces. We worked to address the problem of detecting and isolating collisions between a robotic manipulator and its environment, using only on-board joint torque and position sensing [37]. We presented an algorithm based on a particle filter that, under some assumptions, is able to identify the contact location anywhere on the robot body. It requires the robot to perform small exploratory movements, progressively integrating the new sensing information through a Bayesian framework. The method assumes negligible friction forces, convex contact surfaces, and linear contact stiffness. Compared to existing approaches, it allows this detection to be carried in almost all the surface of the robot's body. We tested the proposed approach both in simulation and in a real environment. Experiments in simulation showed that our approach outperformed two other methods that made simpler assumptions. Experiments in a real environment using a robot with joint torque sensors showed the applicability of the method to real world scenarios and its ability to cope with situations where the algorithm's assumptions did not hold.

6.1.5. Cooperative Localization using Interval Analysis

Participants: Ide Flore Kenmogne Fokam, Vincent Drevelle, Eric Marchand.

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. Quantifier elimination techniques have been used to consider uncertainty in the landmarks positions without adding pessimism in the computed solution. This work has been applied to cooperative localization of UAVs, based on image and range measurements [20].

6.2. Advanced Sensor-Based Control

6.2.1. Sensor-based Trajectory Planning for quadrotor UAVs

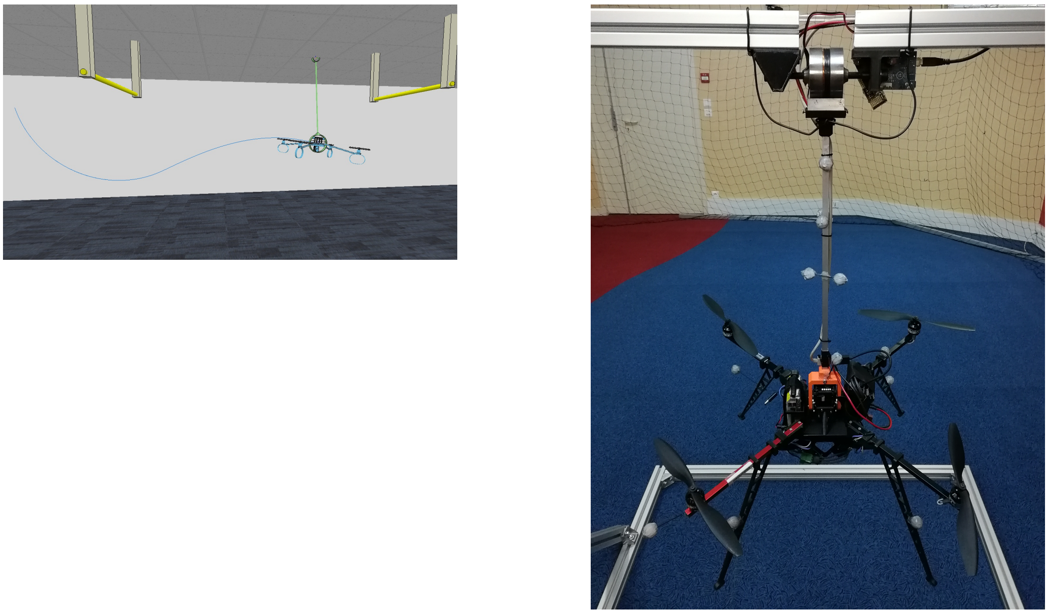
Participants: François Chaumette, Paolo Robuffo Giordano.

In the context of developing robust navigation strategies for quadrotor UAVs with onboard cameras and IMUs, we considered the problem of planning minimum-time trajectors in a cluttered environment for reaching a goal while coping with actuation and sensing constraints [25]. In particular, we considered a realistic model for the onboard camera that considers limited fov and possible occlusions due to obstructed visibility (e.g., presence of obstacles). Whenever the camera can detect landmarks in the environment, the visual cues can be used to drive a state estimation algorithm (a EKF) for updating the current estimation of the UAV state (its pose and velocity). However, beacause of the sensing constraints, the possibility of detecting and tracking the landmarks may be lost while moving in the environment. Therefore, we proposed a robust "perception-aware" planning strategy, based on the bi-directional A^* planner,

6.2.2. 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 "static" contact cases (the UAVs need to exert some forces on the environment in quasi-static 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. During last year we have considered the modeling, control and trajectory planning problem for 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), see Fig. 8(a). We are currently finalizing a real prototype (Fig. 8(b)) for obtaining an experimental validation of the whole approach [1].



(a) (b)

Figure 8. UAVs in Physical Interaction with the Environment. a) The simulated MonkeyRotor performing a hook-to-hook maneuver. b) The prototype currently under finalization.

6.2.3. Trajectory Generation for Minimum Closed-Loop State Sensitivity

Participants: Pascal Brault, Quentin Delamare, Paolo Robuffo Giordano.

The goal of this research activity is to propose a new point of view in addressing the control of robots under parametric uncertainties: rather than striving to design a sophisticated controller with some robustness guarantees for a specific system, we propose to attain robustness (for any choice of the control action) by suitably shaping the reference motion trajectory so as to minimize the *state sensitivity* to parameter uncertainty of the resulting closed-loop system. During this year, we have extended the existing minimization framework to also include the notion of "input sensitivity", which allows to obtain trajectories whose realization (in perturbed conditions) leaves the control inputs unchanged to the largest extent. Such a feature is relevant whenever dealing with, e.g., limited actuation since it guarantees that, even under model perturbations, the inputs do not deviate too much from their nominal values. This novel input sensitivity has been combined with the previously introduced notion of state sensitivity and validated both via monte-carlo simulations and experimentally with a unicycle robot in a large number of tests [1].

6.2.4. Visual Servoing for Steering Simulation Agents

Participants: Axel Lopez Gandia, Eric Marchand, François Chaumette, Julien Pettré.

This research activity is dedicated to the simulation of human locomotion, and more especially to the simulation of the visuomotor loop that controls human locomotion in interaction with the static and moving obstacles of its environment. Our approach is based on the principles of visual servoing for robots. To simulate visual perception, an agent perceives its environment through a virtual camera located in the position of its head. The visual input is processed by each agent in order to extract the relevant information for controlling its motion. In particular, the optical flow is computed to give the agent access to the relative motion of visible objects around it. Some features of the optical flow are finally computed to estimate the risk of collision with obstacle. We have established the mathematical relations between those visual features and the agent's self motion. Therefore, when necessary, the agent motion is controlled and adjusted so as to cancel the visual features indicating a risk of future collision [22], [46].

6.2.5. Strategies for Crowd Simulation Agents

Participants: Wouter Van Toll, Julien Pettré.

This research activity is dedicated to the simulation of crowds based on microscopic approaches. In such approaches, agents move according to local models of interactions that give them the capacity to adjust to the motion of neighbor agents. These purely local rules are not sufficient to produce high-quality long term trajectories through their environment. We provide agents with the capacity to establish mid-term strategies to move through their environment, by establishing a local plan based on their prediction of their surroundings and by verifying regularly this prediction remains valid. In the case validity is not checked, planning a new strategy is triggered [55].

6.2.6. Study of human locomotion to improve robot navigation

Participants: Florian Berton, Julien Bruneau, Julien Pettré.

This research activity is dedicated to the study of human gaze behaviour during locomotion. This activity is directly linked to the previous one on simulation, as human locomotion study results will serve as an input for the design of novel models for simulation. We are interested in the study of the activity of the gaze during locomotion that, in addition to the classical study of kinematics motion parameters, provides information on the nature of visual information acquired by humans to move, and the relative importance of visual elements in their surroundings [36].

6.2.7. Robot-Human Interactions during Locomotion

Participants: Javad Amirian, Fabien Grzeskowiak, Marie Babel, Julien Pettré.

This research activity is dedicated to the design of robot navigation techniques to make them capable of safely moving through a crowd of people. We are following two main research paths. The first one is dedicated to the prediction of crowd motion based on the state of the crowd as sensed by a robot. The second one is dedicated to the creation of a virtual reality platform that enables robots and humans to share a common virtual space where robot control techniques can be tested with no physical risk of harming people, as they remain separated in the physical space. This year, we have delivered techniques for the short term prediction of human locomotion trajectories [34], [35] and robot-human collision avoidance [39].

6.2.8. Visual Servoing for Cable-Driven Parallel Robots

Participant: François Chaumette.

This study is done in collaboration with IRT Jules Verne (Zane Zake, Nicolo Pedemonte) and LS2N (Stéphane Caro) in Nantes (see Section 7.2.2). It is devoted to the analysis of the robustness of visual servoing to modeling and calibration errors for cable-driven parallel robots. The modeling of the closed loop system has been derived, from which a Lyapunov-based stability analysis allowed exhibiting sufficient conditions for ensuring its stability. Experimental results have validated the theoretical results obtained and shown the high robustness of visual servoing for this sort of robots [30], [56].

6.2.9. Visual Exploration of an Indoor Environment

Participants: Benoît Antoniotti, Eric Marchand, François Chaumette.

This study is done in collaboration with the Creative company in Rennes (see Section 6.2.9). It is devoted to the exploration of indoor environments by a mobile robot, Pepper typically (see Section 5.4.2) for a complete and accurate reconstruction of the environment. The exploration strategy we are currently developing is based on maximizing the entropy generated by a robot motion.

6.2.10. Deformation Servoing of Soft Objects

Participant: Alexandre Krupa.

Nowadays robots are mostly used to manipulate rigid objects. Manipulating deformable objects remains challenging due to the difficulty of accurately predicting the object deformations. This year, we developed a model-free deformation servoing method able to do an online estimation of the deformation Jacobian that relates the motion of the robot end-effector to the deformation of a manipulated soft object. The first experimental results are encouraging since they showed that our model-free visual servoing approach based on online estimation provides similar results than a model-based approach based on physics simulation that requires accurate knowledge of the physical properties of the object to deform. This approach has been recently submitted to the ICRA'20 conference.

6.2.11. Multi-Robot Formation Control

Participant: 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. In several previous works we have addressed the problem of coordinating a team of quadrotor UAVs equipped with onboard sensors (such as distance sensors or cameras) for cooperative localization and formation control under the rigidity framework. In [9] an interesting interplay between the rigidity formalism and notions of parallel robotics has been studied, showing how well-known tools from the parallel robotics community can be applied to the multi-robot case, and how these tools can be used for characterizing the stability and singularities of the typical formation control/localization algorithms.

In [17], the problem of distributed leader selection has been addressed by considering agents with a second-order dynamics, thus closer to physical robots that have some unavoidable inertia when moving. This work has extended a previous strategy developed for a first-order case and ported it to the second-order: the proposed algorithm is able to periodically select at runtime the 'best' leader (among the neighbors of the current leader) for maximizing the tracking performance of an external trajectory reference while maintaining a desired formation for the group. The approach has been validated via numerical simulations.

6.2.12. Coupling Force and Vision for Controlling Robot Manipulators

Participants: Alexander Oliva, François Chaumette, Paolo Robuffo Giordano.

The goal of this activity is about coupling visual and force information for advanced manipulation tasks. To this end, we plan to exploit the recently acquired Panda robot (see Sect. 5.4.4), a state-of-the-art 7-dof manipulator arm with torque sensing in the joints, and the possibility to command torques at the joints or forces at the end-effector. Thanks to this new robot, we plan to study how to optimally combine the torque sensing and control strategies that have been developed over the years to also include in the loop the feedback from a vision sensor (a camera). In fact, the use of vision in torque-controlled robot is quite limited because of many issues, among which the difficulty of fusing low-rate images (about 30 Hz) with high-rate torque commands (about 1 kHz), the delays caused by any image processing and tracking algorithms, and the unavoidable occlusions that arise when the end-effector needs to approach an object to be grasped.

Towards this goal, this year we have considered the problem of identification of the dynamical model for the Panda robot [18], by suitably exploiting tools from identification theory. The identified model has been validated in numerous tests on the real robot with very good results and accuracy. A special feature of the model is the inclusion of a (realistic) friction term that accounts well for joint friction (a term that is usually neglected in dynamical model identification).

6.2.13. Subspace-based visual servoing

Participant: Eric Marchand.

To date most of visual servoing approaches have relied on geometric features that have to be tracked and matched in the image. Recent works have highlighted the importance of taking into account the photometric information of the entire images. This leads to direct visual servoing (DVS) approaches. The main disadvantage of DVS is its small convergence domain compared to conventional techniques, which is due to the high non-linearities of the cost function to be minimized. We proposed to project the image on an orthogonal basis (PCA) and then servo on either images reconstructed from this new compact set of coordinates or directly on these coordinates used as visual features [23]. In both cases we derived the analytical formulation of the interaction matrix. We show that these approaches feature a better behavior than the classical photometric visual servoing scheme allowing larger displacements and a satisfactory decrease of the error norm thanks to a well modelled interaction matrix.

6.2.14. Wheelchair Autonomous Navigation for Fall Prevention

Participants: Solenne Fortun, Marie Babel.

The Prisme project (see Section 8.1.4) 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. This year, we finalized the related control framework based on sensor-based servoing principles. We validated the proposed solution through usage tests within the Rehabilitation Center of Pôle Saint Hélier (Rennes).

6.3. Haptic Cueing for Robotic Applications

6.3.1. Wearable Haptics

Participants: Marco Aggravi, Claudio Pacchierotti.

We worked on developing a novel modular wearable finger interface for cutaneous and kinesthetic interaction [11], shown in Fig. 9. It is composed of a 3-DoF fingertip cutaneous device and a 1-DoF finger kinesthetic exoskeleton, which can be either used together as a single device or separately as two different devices. The 3-DoF fingertip device is composed of a static body and a mobile platform. The mobile platform is capable of making and breaking contact with the finger pulp and re-angle to replicate contacts with arbitrarily oriented surfaces.

The 1-DoF finger exoskeleton provides kinesthetic force to the proximal and distal interphalangeal finger articulations using one servo motor grounded on the proximal phalanx. Together with the wearable device, we designed three different position, force, and compliance control schemes. We also carried out three human subjects experiments, enrolling a total of 40 different participants: the first experiment considered a curvature discrimination task, the second one a robotassisted palpation task, and the third one an immersive experience in Virtual Reality. Results showed that providing cutaneous and kinesthetic feedback through our device significantly improved the performance of all the considered tasks. Moreover, although cutaneous-only feedback showed promising performance, adding kinesthetic feedback improved most metrics. Finally, subjects ranked our device as highly wearable, comfortable, and effective.

On the same line of research, this year we guest edited a Special Issue on the IEEE Transactions on Haptics [26]. Thirteen papers on the topic have been published.

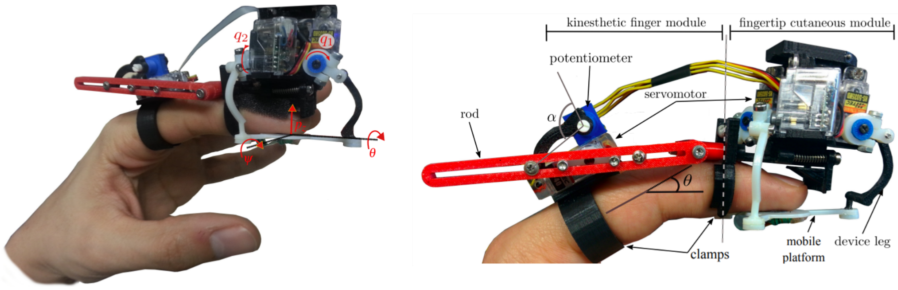


Figure 9. The proposed wearable device. It can provide both cutaneous feedback at the fingertip and kinesthetic feedback at the finger.

6.3.2. Mid-Air Haptic Feedback

Participants: Claudio Pacchierotti, Thomas Howard.

GUIs have been the gold standard for more than 25 years. However, they only support interaction with digital information indirectly (typically using a mouse or pen) and input and output are always separated. Furthermore, GUIs do not leverage our innate human abilities to manipulate and reason with 3D objects. Recently, 3D interfaces and VR headsets use physical objects as surrogates for tangible information, offering limited malleability and haptic feedback (e.g., rumble effects). In the framework of project H-Reality (Sect. 8.3.5), we are working to develop novel mid-air haptics paradigm that can convey the information spectrum of touch sensations in the real world, motivating the need to develop new, natural interaction techniques.

In this respect, we started working on investigating the recognition of local shapes using mid-air ultrasound haptics [45]. We have presented a series of human subject experiments investigating important perceptual aspects related to the rendering of 2D shapes by an ultrasound haptic interface (the Ultrahaptics STRATOS platform). We carried out four user studies aiming at evaluating (i) the absolute detection threshold for a static focal point rendered via amplitude modulation, (ii) the absolute detection and identification thresholds for line patterns rendered via spatiotemporal modulation, (iii) the ability to discriminate different line orientations, and (iv) the ability to perceive virtual bumps and holes.

Our results show that focal point detection thresholds are situated around 560Pa peak acoustic radiation pressure, with no evidence of effects of hand movement on detection. Line patterns rendered through spatiotemporal modulation were detectable at lower pressures, however their shape was generally not recognized as a line below a similar threshold of approx. 540Pa peak acoustic radiation pressure. We did not find any significant effect of line orientation relative to the hand both in terms of detection thresholds and in terms of correct identification of line orientation.

6.3.3. Tangible objects in VR and AR

Participant: Claudio Pacchierotti.

Still in the framework of the H-Reality project (Sect. 8.3.5), we studied the role of employing simple tangible objects in VR and AR scenarios, to improve the illusion of telepresence in these environments. We started by investigating the role of haptic sensations when interacting with tangible objects. Tangible objects are used in Virtual Reality to provide human users with distributed haptic sensations when grasping virtual objects. To achieve a compelling illusion, there should be a good correspondence between the haptic features of the tangible object and those of the corresponding virtual one, i.e., what users see in the virtual environment should match as much as possible what they touch in the real world. For this reason, we aimed at quantifying how similar tangible and virtual objects need to be, in terms of haptic perception, to still feel the same [40]. As it is often not possible to create tangible replicas of all the virtual objects in the scene, it is indeed important to understand how different tangible and virtual objects can be without the user noticing. Of course, the visuohaptic perception of objects encompasses several different dimensions, including the object's size, shape, mass, texture, and temperature. We started by addressing three representative haptic features - width, local orientation, and curvature, - which are particularly relevant for grasping. We evaluated the just-noticeable difference (JND) when grasping, with a thumb-index pinch, a tangible object which differ from a seen virtual one on the above three important haptic features. Results show JND values of 5.75%, 43.8%, and 66.66% of the reference shape for the width, local orientation, and local curvature features, respectively.

As we mentioned above, for achieving a compelling illusion during interaction in VR, there should be a good correspondence between what users see in the virtual environment and what they touch in the real world. The haptic features of the tangible object should – up to a certain extent – match those of the corresponding virtual one. We worked on an innovative approach enabling the use of few tangible objects to render many virtual ones [41]. Toward this objective, we present an algorithm which analyses different tangible and virtual objects to find the grasping strategy best matching the resultant haptic pinching sensation. Starting from the meshes of the considered objects, the algorithm guides users towards the grasping pose which best matches what they see in the virtual scene with what they feel when touching the tangible object. By selecting different grasping positions according to the virtual object to render, it is possible to use few tangible objects to render multiple virtual ones. We tested our approach in a user study. Twelve participants were asked to grasp different virtual objects, all rendered by the same tangible one. For every virtual object, our algorithm found the best pinching match on the tangible one, and guided the participant toward that grasp. Results show that our algorithm was able to well combine several haptically-salient object features to find corresponding pinches between the given tangible and virtual objects. At the end of the experiment, participants were also asked to guess how many tangible objects were used during the experiment. No one guessed that we used only one, proof of a convincing experience.

6.3.4. Wearable haptics for an Augmented Wheelchair Driving Experience

Participants: Louise Devigne, François Pasteau, Marco Aggravi, Claudio Pacchierotti, Marie Babel.

Smart powered wheelchairs can increase mobility and independence for people with disability by providing navigation support. For rehabilitation or learning purposes, it would be of great benefit for wheelchair users to have a better understanding of the surrounding environment while driving. Therefore, a way of providing navigation support is to communicate information through a dedicated and adapted feedback interface.

We then envisaged the use of wearable vibrotactile haptics, i.e. two haptic armbands, each composed of four evenly-spaced vibrotactile actuators. With respect to other available solutions, our approach provides rich navigation information while always leaving the patient in control of the wheelchair motion. We then conducted experiments with volunteers who experienced wheelchair driving in conjunction with the use of the armbands to provide drivers with information either on the presence of obstacles. Results show that providing information on closest obstacle position improved significantly the safety of the driving task (least number of collisions). This work is jointly conducted in the context of ADAPT project (Sect. 8.3.6) and ISI4NAVE associate team (Sect. 8.4.1.1).

6.4. Shared Control Architectures

6.4.1. Shared Control for Remote Manipulation

Participants: Firas Abi Farraj, Paolo Robuffo Giordano, Claudio Pacchierotti, Rahaf Rahal.

As teleoperation systems become more sophisticated and flexible, the environments and applications where they can be employed become less structured and predictable. This desirable evolution toward more challenging robotic tasks requires an increasing degree of training, skills, and concentration from the human operator. For this reason, researchers started to devise innovative approaches to make the control of such systems more effective and intuitive. In this respect, shared control algorithms have been investigated as one the main tools to design complex but intuitive robotic teleoperation system, helping operators in carrying out several increasingly difficult robotic applications, such as assisted vehicle navigation, surgical robotics, brain-computer interface manipulation, rehabilitation. This approach makes it possible to share the available degrees of freedom of the robotic system between the operator and an autonomous controller. The human operator is in charge of imparting high level, intuitive goals to the robotic system; while the autonomous controller translates them into inputs the robotic system can understand. How to implement such division of roles between the human operator and the autonomous controller highly depends on the task, robotic system, and application. Haptic feedback and guidance have been shown to play a significant and promising role in shared control applications. For example, haptic cues can provide the user with information about what the autonomous controller is doing or is planning to do; or haptic force can be used to gradually limit the degrees of freedom available to the human operator, according to the difficulty of the task or the experience of the user. The dynamic nature of haptic guidance enables us to design very flexible robotic system, which can easily and rapidly change the division of roles between the user and autonomous controller.

Along this general line of research, during this year we gave the following contributions:

- in [51] we proposed a shared control algorithm for remote telemanipulation of redundant robots able to fuse the task-prioritized control architecture (for handling the concurrent realization of multiple tasks) with haptic guidance techniques. In particular, we developed a suitable passivity-preserving strategy based on energy tanks for always guaranteeing stability despite the possible presence of autonomous tasks that could generate an increase of energy during operation. The approach has been validated with extensive simulative results in a realistic environment.
- in [6] we have considered a shared control algorithm for telemanipulation that embeds the presence of a grasping planner for guiding the operator towards suitable grasping poses. The operator retains control of the end-effector motion and eventual grasping location, but she/he is assisted by the autonomy (via force cues) in navigating towards good grasps, as classified by the grasping planner that takes as input a RGBD image of the scene and computes a set of grasping poses along the object contour.
- in [50] we have presented two haptic shared-control approaches for robotic cutting. They are designed to assist the human operator by enforcing different nonholonomic-like constraints representative of the cutting kinematics. To validate this approach, we carried out a human-subject experiment in a real cutting scenario. We compared our shared-control techniques with each other and with a standard haptic teleoperation scheme. Results show the usefulness of assisted control schemes in complex applications such as cutting.

6.4.2. Teleoperation of Flexible Needle with Haptic Feedback and Ultrasound Guidance

Participants: Jason Chevrie, Alexandre Krupa, Marie Babel.

Needle insertion procedures under ultrasound guidance are commonly used for diagnosis and therapy. This kind of intervention can greatly benefit from robotic systems to improve their accuracy and success rate. In the past years, we have developed a robotic framework dedicated to 3D steering of beveled-tip flexible needle in order to autonomously reach a desired target in the tissues by ultrasound visual servoing using a 3D ultrasound probe. This year we have proposed a real-time semi-automatic teleoperation framework that enables the user to directly control the trajectory of the needle tip during its insertion via a haptic interface [38]. The framework enables the user to intuitively guide the trajectory of the needle tip in the ultrasound 3D volume while the controller handles the complexity of the 6D motion that needs to be applied to the needle base. A mean targeting accuracy of 2.5 mm has been achieved in gelatin phantoms and different ways to provide the haptic feedback as well as different levels of control given to the user on the tip trajectory have been compared. Limiting the user input to the insertion speed while automatically controlling the trajectory of the needle tip seems to provide a safer insertion process, however it may be too constraining and can not handle situations where more control over the tip trajectory is required, for example if unpredicted obstacles need to be avoided. On the contrary, giving the full control of the 3D tip velocity to the user and applying a haptic feedback to guide the user toward the target proved to maintain a low level of needle bending and tissue deformation.

6.4.3. Needle Comanipulation with Haptic Guidance

Participants: Hadrien Gurnel, Alexandre Krupa.

The objective of this work is to provide assistance during manual needle steering for biopsies or therapy purposes (see Section 7.2.3). At the difference of our work presented in Section 6.4.2 where a robotic system is used to steer the needle, we propose in this study another way of assistance where the needle is collaboratively manipulated by the physician and a haptic device. The principle of our approach is to provide haptic cue feedback to the clinician in order to help him during his manual gesture [43]. We elaborated 5 different haptic-guidance strategies to assist the needle pre-positioning and pre-orienting on a pre-defined insertion point, and with a pre-planned desired incidence angle. The haptic guides rely on the position and orientation errors between the needle, the entry point and the desired angle of incidence toward the target, which are computed from the measurements provided by an electromagnetic tracker. Each of the guide implements a different Guiding Virtual Fixture producing haptic cues that attract the needle towards a point or a trajectory in space with different force feedback applied on the user's hand manipulating the needle. A two-step evaluation was conducted to assess the performance and ergonomy of each haptic guide, and compare them to the unassisted reference gesture. The first evaluation stage [44] involved two physicians both experts in needle manipulation at Rennes University Hospital. The performance results showed that, compared to the unassisted gesture, the positioning accuracy was enhanced with haptic guidance. The second evaluation stage [43] was a user study with twelve participants. From this second study it results that the most constraining guide allows to perform the gesture with the best accuracy, lower time duration and highest level of ergonomy.

6.4.4. Shared Control of a Wheelchair for Navigation Assistance

Participants: Louise Devigne, Marie Babel.

Power wheelchairs allow people with motor disabilities to have more mobility and independence. However, driving safely such a vehicle is a daily challenge particularly in urban environments while navigating on sidewalks, negotiating curbs or dealing with uneven grounds. Indeed, differences of elevation have been reported to be one of the most challenging environmental barrier to negotiate, with tipping and falling being the most common accidents power wheelchair users encounter. It is thus our challenge to design assistive solutions for power wheelchair navigation in order to improve safety while navigating in such environments. To this aim, we proposed a shared-control algorithm which provides assistance while navigating with a wheelchair in an environment consisting of negative obstacles. We designed a dedicated sensor-based control law allowing trajectory correction while approaching negative obstacles e.g. steps, curbs, descending slopes. This shared control method takes into account the humanin-the loop factor. In this study, our solution the ability of our system to ensure a safe trajectory while navigating on a sidewalk is demonstrated through simulation, thus providing a proof-of-concept of our method [42].

6.4.5. Wheelchair-Human Interactions during crossing situations

Participants: Marie Babel, Julien Pettré.

Designing smart powered wheelchairs requires to better understand interactions between walkers and such vehicles. We focus on collision avoidance task between a power wheelchair (fully operated by a human) and a walker, where the difference in the nature of the agents (weight, maximal speed, acceleration profiles) results into asymmetrical physical risk in case of a collision, for example due to the protection power wheelchair provides to its driver, or the higher energy transferred to the walker during head-on collision.

We then conducted experiments with Results show that walkers set more conservative strategies when interacting with a power wheelchair. These results can then be linked to the difference in the physical characteristics of the walkers and power wheelchairs where asymmetry in the physical risks raised by collisions influence the strategies performed by the walkers in comparison with a similar walker-walker situation. This gives interesting insights in the task of modeling such interactions, indicating that geometrical terms are not sufficient to explain behaviours, physical terms linked to collision momentum should also be considered [49][62].

6.4.6. *Multisensory power wheelchair simulator*

Participants: Guillaume Vailland, Louise Devigne, François Pasteau, Marie Babel.

Power wheelchair driving is a challenging task which requires good visual, cognitive and visuo-spatial abilities. Besides, a power wheelchair can cause material damage or represent a danger of injury for others or oneself if not operated safely. Therefore, training and repeated practice are mandatory to acquire safe driving skills to obtain power wheelchair prescription from therapists. However, conventional training programs may reveal themselves insufficient for some people with severe impairments. In this context, Virtual Reality offers the opportunity to design innovative learning and training programs while providing realistic wheelchair driving experience within a virtual environment. We then proposed a user-centered design of a multisensory power wheelchair simulator [59][58]. This simulator addresses classical virtual experience drawbacks such as cybersickness and sense of presence by combining 3D visual rendering, haptic and vestibular feedback. It relies on a modular and versatile workflow enabling not only easy interfacing with any virtual display, but also with any user interface such as wheelchair controllers or feedback devices. First experiments with able-bodied people shown that vestibular feedback activation increases the Sense of Presence and decreases cybersickness [54].

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. *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 Rainbow medical robotic platform (see Sect. 5.4.3). Each module is rent 40 days during a 3-year period in the context of the IRT B<>com NeedleWare project (see Section 7.2.3).

7.2. Bilateral Grants with Industry

7.2.1. *Creative*

Participants: Benoît Antoniotti, François Chaumette, Eric Marchand.

No Inria Rennes 13996, duration: 36 months.

This project funded by Creative started in March 2019. It supports Benoît Antoniotti's Ph.D. about visual exploration (see Section 6.2.9).

7.2.2. *IRT JV Perform*

Participant: François Chaumette.

No Inria Rennes 14049, duration: 36 months.

This project funded by IRT Jules Verne in Nantes started in January 2018. It is achieved in cooperation with Stéphane Caro from LS2N in Nantes to support Zane Zake's Ph.D. about visual servoing of cable-driven parallel robots (see Section 6.2.8).

7.2.3. *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 visual control for assisting manual steering of needles for biopsy or therapy purposes in a synergetic way (see Section 6.4.3). This year, we published [43] [44] in the scope of this project.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. SAD WH-DRONE

Participants: Marco Aggravi, Claudio Pacchierotti.

no CNRS Rennes 181089, duration: 24 months.

This project funded by the Brittany council started in January 2019. It supports in part Marco Aggravi's research on using wearable interfaces for flying swarms of drones.

8.1.2. 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 supports the activities related to the teleoperation of drones (quadrotor UAVs) using wearable haptics interfaces.

8.1.3. IRT Jules Verne Happy

Participant: François Chaumette.

no Inria Rennes 13521, duration: 36 months.

This project started in June 2018. It is managed by IRT Jules Verne in Nantes and achieved in cooperation with LS2N and Airbus. Its goal is to develop local sensor-based control methods for the assembly of large parts of aircrafts.

8.1.4. Prisme

Participants: Solenne Fortun, François Pasteau, Marie Babel.

no Insa Rennes 2017-0004, duration: 36 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 addressed (see Section 6.2.14).

8.1.5. Silver Connect

Participant: Marie Babel.

no Insa Rennes 2018-0076, duration: 34 months.

This project started in November 2018 and is supported by Brittany region/BPI as well as FEDER. This project aims at designing a fall detection framework by means of vision-based algorithms coupled with deep learning solutions.

8.1.6. Cartam

Participants: Noura Neji, Fabien Spindler, François Chaumette.

no Inria 13954 and 14041, duration: 36 months.

This project started in January 2019 and is supported by Brittany region and FEDER. It is managed by Triskalia with Unilet, Copeeks, Neotec Vision, Rainbow group, and our start-up Dilepix. It aims at designing a vision system able to detect adventices in a field. We are in charge of tracking the adventices once they are detected and of building a mosaic of the field for locating them.

8.2. National Initiatives

8.2.1. ANR PLaTINUM

Participant: Vincent Drevelle.

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 Inria (Chorale group in Sophia-Antipolis and Rainbow). The project is focused on robust long-term mapping of urban environments. Map building consists in the acquisition of a textured 3-D model of urban environment, and automatic semantic labelling of the environment features (roads, buildings, cars, etc.). From this model, an optimal representation is generated and made available in the cloud, in the form of a network of RGB-D-L spheres storing photometry, geometry (depth) and object labels. Mobile agents are able to determine their position and navigate in the sphere graph using dense matching. Agents upload significant environment changes to the sphere server in cloud, for map update purposes.

8.2.2. ANR Sesame

Participant: François Chaumette.

no Inria 13722, duration: 48 months.

This project started in January 2019. It involves a consortium managed by LS2N (Nantes) with LIP6 (Paris) and Rainbow group. It aims at analysing singularity and stability issues in visual servoing.

8.2.3. Equipex Robotex

Participants: Fabien Spindler, François Chaumette.

no Inria Rennes 6388, duration: 9 years.

Rainbow 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 obtained the humanoid robot Romeo in 2015.

8.3. European Initiatives

FP7 & H2020 Projects

8.3.1. FP7 Space RemoveDEBRIS

Participants: Eric Marchand, François Chaumette.

Instrument: Specific Targeted Research Project

Duration: October 2013 - March 2019

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: A huge amount of debris have progressively been generated since the beginning of the space era. Most of the objects launched into space are still orbiting the Earth and today these objects and their by-products represent a threat both in space and on Earth. In Space, debris lead to collisions and therefore to damages to operational satellites. For both issues, a credible solution has emerged over the recent years: actively removing heavy debris objects by capturing them and then either disposing them by destructive re-entry in Earth atmosphere or disposing them in graveyard orbits. The RemoveDEBRIS project aimed to demonstrate key technologies for ADR in three main domains by performing in-orbit demonstrations representative of an ADR mission. The specific key technologies that have been demonstrated as part of this project are: (i) Capture technologies such as nets and harpoons (ii) De-orbiting technologies such as electric propulsion and drag augmentation (iii) Proximity Rendezvous operations technologies based on vision-based navigation. The technology demonstrations has been carried in orbit using a micro satellite test-bed, a world's first. The micro satellite has carried the ADR payloads together with two deployable nanosatellites (CubeSats). Through a series of operations, the nanosatellites have been ejected, re-captured, inspected and de-orbited, thereby demonstrating the ADR key technologies [16], [8], [7]. Our goal in this long project was to develop and validate model-based tracking algorithms on images acquired during the actual space debris removal mission [47].

8.3.2. H2020 ICT Comanoid

Participants: Fabien Spindler, François Chaumette.

Title: Multi-contact Collaborative Humanoids in Aircraft Manufacturing

Programme: H2020

Duration: January 2015 - February 2019

Coordinator: CNRS (Lirmm)

Partners: Airbus Group (France), DLR (Germany), Università Degli Studi di Roma La Sapienza (Italy), CNRS (I3S)

Inria contact: François Chaumette

Abstract: Comanoid investigated 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 was 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 suggested 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 aimed at assessing clearly how far the state-of-the-art stands from such novel technologies. In particular the project focused on implementing a real-world humanoid robotics solution using the best of research and innovation. The main challenge was 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 [21].

8.3.3. H2020 ICT CrowdBot

Participants: Javad Amirian, Fabien Grzeskowiak, Solenne Fortun, Marie Babel, Julien Pettré, Fabien Spindler.

Title: Robot navigation in dense crowds

Programme: H2020

Duration: Jan 2018 - Jun 2021

Coordinator: Inria

Partners: UCL (UK), SoftBank Robotics (France), Univ. Aachen (Germany), EPFL (Switzerland), ETHZ (Switzerland), Locomotec (Germany)

Inria contact: Julien Pettré

Abstract: CROWDBOT will enable mobile robots to navigate autonomously and assist humans in crowded areas. Today's robots are programmed to stop when a human, or any obstacle is too close, to avoid coming into contact while moving. This prevents robots from entering densely frequented areas and performing effectively in these high dynamic environments. CROWDBOT aims to fill in the gap in knowledge on close interactions between robots and humans during navigation tasks. The project considers three realistic scenarios: 1) a semi-autonomous wheelchair that must adapt its trajectory to unexpected movements of people in its vicinity to ensure neither its user nor the pedestrians around it are injured; 2) the commercially available Pepper robot that must navigate in a dense crowd while actively approaching people to assist them; 3) the under development robot cuyBot will adapt to compact crowd, being touched and pushed by people. These scenarios generate numerous ethical and safety concerns which this project addresses through a dedicated Ethical and Safety Advisory Board that will design guidelines for robots engaging in interaction in crowded environments. CROWDBOT gathers the required expertise to develop new robot capabilities to allow robots to move in a safe and socially acceptable manner. This requires achieving step changes in a) sensing abilities to estimate the crowd motion around the robot, b) cognitive abilities for the robot to predict the short term evolution of the crowd state and c) navigation abilities to perform safe motion at close range from people. Through demonstrators and open software components, CROWDBOT will show that safe navigation tasks can be achieved within crowds and will facilitate incorporating its results into mobile robots, with significant scientific and industrial impact. By extending the robot operation field toward crowded environments, we enable possibilities for new applications, such as robot-assisted crowd traffic management.

8.3.4. H2020 ICT PRESENT

Participants: Adèle Colas, Alberto Jovane, Claudio Pacchierotti, Julien Pettré.

Title: Photoreal REaltime Sentient ENTity

Programme: H2020

Duration: Sep 2019 - Aug 2022

Coordinator: Univ Pompeu Fabra (Spain)

Partners: The Framestore Ltd (UK), Cubic Motion Ltd (UK), InfoCert Spa (Italy), Brainstorm Multimedia S.L. (ES), Creative Workers - Creatieve Werkers VZW (Belgium), Universitaet Augsburg (Germany), Inria (France)

Inria contact: Julien Pettré

Abstract: PRESENT is a three-year Research and Innovation project to create virtual digital companions—embodied agents—that look entirely naturalistic, demonstrate emotional sensitivity, can establish meaningful dialogue, add sense to the experience, and act as trustworthy guardians and guides in the interfaces for AR, VR and more traditional forms of media.

There is no higher quality interaction than the human experience when we use all our senses together with language and cognition to understand our surroundings and—above all—to interact with other people. We interact with today's Ïntelligent Personal Assistantsprimarily by voice; communication is episodic, based on a request-response model. The user does not see the assistant, which does not take advantage of visual and emotional clues or evolve over time. However, advances in the real-time creation of photorealistic computer generated characters, coupled with emotion recognition and behaviour, and natural language technologies, allow us to envisage virtual agents that are realistic in both looks and behaviour; that can interact with users through vision, sound, touch and movement as they navigate rich and complex environments; converse in a natural manner; respond to moods and emotional states; and evolve in response to user behaviour.

PRESENT will create and demonstrate a set of practical tools, a pipeline and APIs for creating realistic embodied agents and incorporating them in interfaces for a wide range of applications in entertainment, media and advertising.

8.3.5. H2020 FET-OPEN H-Reality

Participants: Claudio Pacchierotti, Paolo Robuffo Giordano, François Chaumette.

Title: Mixed Haptic Feedback for Mid-Air Interactions in Virtual and Augmented Realities

Programme: H2020

Duration: October 2018 - September 2021

Coordinator: Univ. Birmingham (UK)

Partners: Univ. Birmingham (UK, coordinator), CNRS (France), TU Delft (NL), Ultrahaptics (UK) and Actronika SAS (France)

CNRS contact: Claudio Pacchierotti

Abstract: Digital content today remains focused on visual and auditory stimulation. Even in the realm of VR and AR, sight and sound remain paramount. In contrast, methods for delivering haptic (sense of touch) feedback in commercial media are significantly less advanced than graphical and auditory feedback. Yet without a sense of touch, experiences ultimately feel hollow, virtual realities feel false, and Human-Computer Interfaces become unintuitive. Our vision is to be the first to imbue virtual objects with a physical presence, providing a revolutionary, untethered, virtual-haptic reality: H-Reality. The ambition of H-Reality will be achieved by integrating the commercial pioneers of ultrasonic “non-contact” haptics, state-of-the-art vibrotactile actuators, novel mathematical and tribological modelling of the skin and mechanics of touch, and experts in the psychophysical rendering of sensation. The result will be a sensory experience where digital 3D shapes and textures are made manifest in real space via modulated, focused, ultrasound, ready for the untethered hand to feel, where next-generation wearable haptic rings provide directional vibrotactile stimulation, informing users of an object’s dynamics, and where computational renderings of specific materials can be distinguished via their surface properties. The implications of this technology will be far-reaching. The computer touch-screen will be brought into the third dimension so that swipe gestures will be augmented with instinctive rotational gestures, allowing intuitive manipulation of 3D data sets and strolling about the desktop as a virtual landscape of icons, apps and files. H-Reality will transform online interactions; dangerous machinery will be operated virtually from the safety of the home, and surgeons will hone their skills on thin air. Rainbow is involved in H-Reality in cooperation with Anatole Lécuyer and Maud Marchal from the Hybrid group.

Collaborations in European Programs, Except FP7 & H2020

8.3.6. 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: Jan 2017 - Jun 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), 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.

Collaborations with Major European Organizations

8.3.7. *ANR Opmops*

Participants: Florian Berton, Julen Bruneau, 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 that 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.

8.3.8. *iProcess*

Participants: Agniva Sengupta, François Chaumette, Alexandre Krupa, Eric Marchand, Fabien Spindler.

Project acronym: i-Process

Project title: Innovative and Flexible Food Processing Technology in Norway

Duration: January 2016 - December 2019

Coordinator: Sintef Ocean (Norway)

Other partners: Nofima, Univ. of Stavanger, NMBU, NTNU (Norway), DTU (Denmark), KU Leuven (Belgium), and about 10 Norwegian companies.

Abstract: This project was granted by the Norwegian Government. Its main objective was to develop novel concepts and methods for flexible and sustainable food processing in Norway. In the scope of this project, the Rainbow group was involved for visual tracking and visual servoing of generic and potentially deformable objects (see Section 6.1.1 and Section 6.1.2). This year, we published [52], [53] in the scope of this project.

8.3.9. GentleMAN

Participants: Alexandre Krupa, Eric Marchand, François Chaumette, Fabien Spindler.

Project acronym: GentleMAN

Project title: Gentle and Advanced Robotic Manipulation of 3D Compliant Objects

Duration: August 2019 - December 2023

Coordinator: Sintef Ocean (Norway)

Other partners: NTNU (Norway), NMBU (Norway), MIT (USA) and QUT (Australia).

Abstract: This project is funded by the Norwegian Government. Its main objective is to develop a novel learning framework that uses visual, force and tactile sensing to develop new multi-modal learning models, interfaced with underlying robot control, for enabling robots to learn new and advanced skills for the manipulation of 3D compliant objects. In the scope of this project, the Rainbow group is involved in the elaboration of new approaches for visual tracking of deformable objects, active vision and visual servoing for deforming soft objects into desired shapes.

8.4. International Initiatives

8.4.1. Inria Associate Teams Not Involved in an Inria International Labs

8.4.1.1. 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: 2019

See also: https://team.inria.fr/isi4nave/

Using a wheelchair allows people with disability to compensate a loss of mobility. However only 5 to 15% of the 70 million people worldwide who require a wheelchair have access to this type of technical aid. In particular, visual, visuo-spatial and/or cognitive impairments can alter the ability of an individual to independently operate a wheelchair safely.

This project focuses then on two main complementary objectives:

1. to compensate both sensorimotor disabilities and cognitive impairments by designing adapted interfaces,
2. to enhance the driving experience and to bring a new tool for rehabilitation purposes by defining efficient physical Human-Robot Interaction.

In order to ensure a widespread use of robotic systems, innovative interfaces, enabling relevant feedback (medically validated), constitute a major challenge. Trajectory corrections, obtained thanks to an assistance module, will have to be perceived by the user by means of sensitive (visual, tactile…) feedback that will have to be easily adapted to the pathology. Conversely, user interaction with the robotic system can be interpreted to control the wheelchair. Designing such systems require a multidisciplinary study, including medical data collection and analysis.

In our preliminary works, we demonstrated the relevance of share control frameworks. The scope of this new ISI4NAVE Associate Team is then to provide advanced and innovative solutions for controlling wheelchair as well as providing appropriate and relevant feedback to users.

8.4.2. Participation in Other International Programs

8.4.2.1. ACRV

François Chaumette is one of the five external experts 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, Alexander Oliva received a grant to participate to the 2019 Robotic Vision Summer School in Kioloa (New South Wales) and spent a 1-week visit at QUT in March 2019.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Organisation

9.1.1.1. General Chair, Scientific Chair

- Marie Babel was the Scientific Chair and the General co-chair of the workshop “Robotique de la marche” organized in Inria Rennes on December 13th 2019.

9.1.1.2. Member of the Organizing Committees

- P. Robuffo Giordano co-organized with F. Schiano and D. Floreano (EPFL) the Workshop on “Aerial Swarms” at the 2019 International Conference on Intelligent Systems and Robots, Macau, China

9.1.2. Scientific Events: Selection

9.1.2.1. Chair of Conference Program Committees

- Paolo Robuffo Giordano: Program Co-Chair of MRS 2019
- François Chaumette: Program Co-Chair of ICRA 2020 to be held in Paris in June 2020.

9.1.2.2. Member of the Conference Program Committees

- Claudio Pacchierotti: Eurohaptics (Associate Editor), World Haptics (Associate Editor), ICRA 2020 (Associate Editor), Demonstrations Co-Chair, IEEE Haptics Symposium (HAPTICS) 2020
- Paolo Robuffo Giordano: ICRA 2020 (Associate Editor), RSS 2019 (Area Chair)
- Eric Marchand: ICRA 2020 (Associate Editor)
- Julien Pettré: IEEE Virtual Reality 2019, ACM Motion and Interactions in Games 2019, ACM Symposium on Computer Animation

9.1.2.3. Reviewer

- Alexandre Krupa: IROS 2019 (1), ICRA 2020 (1)
- Eric Marchand: IROS 2019 (1), ICRA 2019 (2), ICRA 2020 (1)
- Marie Babel: ICRA 2020 (4), SMC 2019 (1)
- Paolo Robuffo Giordano: ICRA 2020 (2)
- François Chaumette: ICRA 2020 (1), CDC 2019 (1), RoMoCo 2019 (1)

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

- Paolo Robuffo Giordano is Editor for the IEEE Transactions on Robotics since January 2019
- Alexandre Krupa was Associate Editor of IEEE Robotics and Automation Letters (RA-L) (until June 2019) and is currently Associate Editor for the IEEE Transactions on Robotics (since July 2019)
- Eric Marchand is a Senior Editor of IEEE Robotics and Automation Letters (RA-L) and an Editor for Interstices)i(

- François Chaumette: Editor of the IEEE Transactions on Robotics, Editorial Board of the Int. Journal of Robotics Research, Editorial Board of the Springer Tracts in Advanced Robotics, Board Member of the Springer Encyclopedia of Robotics.
- Julien Pettré is Associate Editor for the Wiley's Computer Graphics Forum, and Associate Editor for the Wiley's Computer Animation and Virtual Worlds

9.1.3.2. Reviewer - Reviewing Activities

- Vincent Drevelle: ActaCybernetica
- Eric Marchand: IEEE RA-L (2)
- Marie Babel: IEEE RA-L (1)
- Paolo Robuffo Giordano: IJRR (1), IEEE RA-L (1), IEEE T-CST (2), IEEE T-RO (2)
- François Chaumette: IEEE Trans. on Industrial Electronics (2), IEEE RA-L (1), Acta Astronautica (1)
- Claudio Pacchierotti: RAL (1), IJHCS (5), TOH (7), JMRR (1), ICRA (4), TIE (2), TVCG (1), Frontiers in Robotics and AI (2), THMS (1), Presence (1), Scientific Reports (1)

9.1.4. Invited Talks

- Paolo Robuffo Giordano:
  - "Human-assisted Robotics", Robotics and Mechatronics lab, University of Twente, Netherlands, October 2019
  - "Collective Control, State Estimation and Human Interaction for Quadrotors in Unstructured Environments", Technoférence, Images et Réseaux, Rennes, October 2019
  - "Human-assisted Robotics", Robotics Research Jam Session, Univ. of Pisa, Italy, April 2019
- Claudio Pacchierotti:
  - with A. M. Okamura, "Learnings from cross-functional work on skin stretch haptics." Workshop "Cross-functional collaboration between engineering & perception researchers" at WHC Conference, Tokyo, Japan, 2019.
  - "Cutaneous haptic technologies for robotics and immersive environments." Cirrus Logic, Edinburgh, United Kingdom, 2019.
  - "Virtual Reality and Robotics for Industry 4.0." University of Aarhus, Herning, Denmark, 2019.
- Marie Babel:
  - "Simulation virtuelle et retour haptique du mouvement : une expérience de conduite réaliste immersive", Séminaire d'hiver de l'IFRATH, Paris, January 2019
  - "Fauteuil roulant intelligent : croiser les regards (académiques et cliniques)", Technoférence, Images et Reseaux, January 2019
- François Chaumette:
  - Plenary talk on "Geometric and end-to-end robot vision-based control" at 12th Int. Workshop on Robot Motion and Control (RoMoCo 2019), Poznan, Poland, July 2019 [33].
  - "Sensor-based Control", GdR Robotics Winter School : Robotics Principia, Sophia-Antipolis, January 2019.
- Fabien Spindler:
  - "Avancées en vision pour la saisie et la manipulation", Journée GT3 du GDR robotique, ISIR, Paris, Juin 2019

9.1.5. Leadership within the Scientific Community

- Claudio Pacchierotti is the Chair of the IEEE Technical Committee on Haptics and the Secretary of the Eurohaptics Society.
- François Chaumette is a member of the Scientific Council of RTE Vedecom, ImVIA lab in Dijon and Le Creusot, and a founding member of the Scientific Council of the “GdR Robotique”.

9.1.6. Scientific Expertise

- Paolo Robuffo Giordano was reviewer for evaluating proposals of the ANR (French National Research Agency), Mitacs (Canada), SNSF (Swiss National Science Foundation), MIUR (Italian National Research Agency), and the EU commission (ERC Consolidator Grants)
- Claudio Pacchierotti was reviewer for evaluating proposals of the European Science Foundation (ESF).
- Fabien Spindler was a member of the HCERES evaluation committee of URIA, the laboratory of the Ecole des Mines de Douai, the DTIS laboratory of Onera and the LISTIC in Annecy.
- Marie Babel serves as an expert for the International Mission of the French Research Ministry (MEIRIES) - Campus France (about 30 expertises per year). She was also reviewer for evaluating proposals of the ANR (French National Research Agency)
- François Chaumette served for the fourth (and last) time as Panel Member of the PE7 ERC Consolidator Grants. He also served as member of IEEE RAS Pioneer Award Evaluation Panel and IEEE RAS Awards Evaluation Panel. He was a panel member of the 2019 ICREA Academia program (similar to IUF for Catalonia). Finally, he served in the jury for the Inria research positions (CR2) available for the Inria Rennes-Bretagne Atlantique centre.

9.1.7. Research Administration

- Alexandre Krupa is a member of the CUMIR (“Commission des Utilisateurs des Moyens Informatiques pour la Recherche”) of Inria Rennes-Bretagne Atlantique, he also serves as Inria representative (correspondent) at the IRT Jules Verne.
- Eric Marchand served as secretary in the board of the “Association Française pour la Reconnaissance et l’interprétation des Formes” (AFRIF). He is the head of "Digital Signals and Images, Robotics" department at IRISA.
- 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, and of the Scientific Steering Committee (COSS) of IRISA. Since December 2019, he is a member of the Inria COERLE committee (in charge of the ethical aspects of all Inria research).

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Marie Babel:

Master INSA2: “Robotics”, 26 hours, M1, INSA Rennes

Master INSA1: “Concepts de la logique à la programmation”, 20 hours, L3, INSA Rennes

Master INSA1: “Langage C”, 12 hours, L3, INSA Rennes

Master INSA2: “Computer science project”, 30 hours, M1, INSA Rennes

Master INSA1: “Practical studies”, 16 hours, L3, INSA Rennes

Master INSA2: “Image analysis”, 26 hours, M1, INSA Rennes

Master INSA1: “Remedial math courses”, 50 hours, L3, INSA Rennes

François Chaumette:

Master SISEA: “Robot Vision”, 12 hours, M2, Université de Rennes 1

Master ENS: “Visual servoing”, 6 hours, M1, Ecole Nationale Supérieure de Rennes

Master ESIR3: “Visual servoing”, 8 hours, M2, Ecole supérieure d’ingénieurs de Rennes

Vincent Drevelle:

Master ILA: “Terrain information systems”, 30 hours, M2, Université de Rennes 1

Master Info: “Artificial intelligence”, 20 hours, M1, Université de Rennes 1

Licence Info: “Computer systems architecture”, 42 hours, L1, Université de Rennes 1

Master Elec: “Electronics project”, 16 hours, M1, Université de Rennes 1

Portail Info-Elec: “Discovering programming and electronics”, 22 hours, L1, Université de Rennes 1

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: “Multisensor data fusion”, 20 hours, M2, Université de Rennes 1

Master IL: “Mobile robotics”, 32 hours, M2, Université de Rennes 1

Alexandre Krupa:

Master FIP TIC-Santé: “Ultrasound visual servoing”, 6 hours, M2, Télécom Physique Strasbourg

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 ENS: “Computer vision”, 16 hours, M2, ENS Rennes

Master MIA: “Augmented reality”, 4 hours, M2, Université de Rennes 1

Julien Pettré:

Master SIF: "Motion for Animation and Robotics", 6 hours, Unviersité de Rennes 1

Master Artificial Intelligence and Advanced Visual Computing: "Advanced 3D graphics", 3 hours, Ecole Polytechnique

9.2.2. Supervision

- Ph.D. in progress: Agniva Sengupta, “Visual tracking of deformable objects with RGB-D camera”, started in January 2017, supervised by Alexandre Krupa and Eric Marchand
- Ph.D. in progress: Xavier De Tinguy de la Girouliére, “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: Romain Lagneau (Hybrid group), “Control of soft object deformation by visual servoing”, started in October 2017, supervised by Alexandre Krupa and Maud Marchal (Hybrid group)

- Ph.D. in progress: Florian Berton, “Gaze analysis for crowd behaviours study”, started in October 2017, supervised by Anne-Hélène Olivier (MimeTIC group), Ludovic Hoyet (MimeTIC group) and Julien Pettré
- Ph.D. in progress: Zane Zake (IRT Jules Verne), “Visual servoing for cable-driven parallel robots”, started in January 2018, supervised by Stéphane Caro (LS2N) and François Chaumette
- Ph.D. in progress: Javad Amirian, “Crowd motion prediction for robot navigation in dense crowds”, started in January 2018, supervised by Jean-Bernard Hayet (CIMAT, Guana) and Julien Pettré
- Ph.D. in progress: Xi Wang “Robustness of Visual SLAM techniques to light changing conditions”, started in September 2018, supervised by Eric Marchand and Marc Christie (MimeTIC group)
- Ph.D. in progress: Alexander Oliva, “Coupling Vision and Force for Robotic Manipulation”, started in October 2018, supervised by François Chaumette and Paolo Robuffo Giordano
- Ph.D. in progress: Ketty Favre “Lidar-based localization”, started in October 2018, supervised by Eric Marchand, Muriel Pressigout and Luce Morin (IMAGE group)
- Ph.D. in progress: Fabien Grzeskowiak, “Crowd simulation for testing robot navigation in dense crowds”, started in October 2018, supervised by Marie Babel and Julien Pettré
- Ph.D. in progress: Guillaume Vailland, “Outdoor wheelchair assisted navigation: reality versus virtuality”, started in November 2018, supervised by Marie Babel and Valérie Gouranton (Hybrid group)
- Ph.D. in progress: Benoît Antoniotti “Scene reconstruction and exploration”, started in March 2019, supervised by Eric Marchand, François Chaumette, and Antoine Lehuger (Creative)
- Ph.D. in progress: Pascal Brault, “Planification et optimisation de trajectoires robustes aux incertitudes paramétriques pour des taches robotiques fondées sur l’usage de capteurs”, started in September 2019, supervised by Paolo Robuffo Giordano and Quentin Delamare
- Ph.D. in progress: Alberto Jovane, “Modélisation de mouvements réactifs et comportements non verbaux pour la création d’acteurs digitaux pour la réalité virtuelle”, started in September 2019, supervised by Marc Christie (MimeTIC group), Claudio Pacchierotti, Ludovic Hoyet (MimeTIC group) and Julien Pettré
- Ph.D. in progress: Samuel Felton “Deep Learning for visual servoing”, started in October 2019, supervised by Eric Marchand and Elisa Fromont (Lacodam group)
- Ph.D. in progress: Mathieu Gonzalez “SLAM in time varying environment”, started in October 2019, supervised by Eric Marchand and Jérome Royan (IRT B<>COM)
- Ph.D. in progress: Adèle Colas, “Modélisation de comportements collectifs réactifs et expressifs pour la réalité virtuelle”, started in November 2019, supervised by Claudio Pacchierotti, Anne-Hélène Olivier, Ludovic Hoyet (MimeTIC group) and Julien Pettré
- Ph.D. defended: Salma Jiddi, “Photometric Registration of Indoor Real Scenes using an RGB-D Camera with Application to Mixed Reality", defended in January 2019, supervised by Eric Marchand [2]
- Ph.D. defended: Ide-Flore Kenmogne “Localisation ensembliste de drones à l’aide de méthodes par intervalles”, defended in March 2019, supervised by Eric Marchand and Vincent Drevelle [3]
- Ph.D. defended: Quentin Delamare, “Algorithmes d’estimation et de commande pour des quadrirotors en interaction physique avec l’environnement”, defended on December 2019, supervised by Paolo Robuffo Giordano and Antonio Franchi (LAAS) [1]
- Ph.D. defended: Axel Lopez Gandia, “Optical flow-based navigation algorithms for virtual humans”, defended on December 2019, supervised by Julien Pettré and François Chaumette [4]
- Ph.D. to be defended: Hadrien Gurnel “Needle comanipulation with haptic guidance for percutaneous interventions”, to be defended in January 2020, supervised by Alexandre Krupa, Maud Marchal (Hybrid group) and Laurent Launay (IRT B<>com)

9.2.3. Internships

- Alexandre Bonneau, ENS Rennes - Univ. Rennes 1, “Haptic rendering of virtual contacts in crowded environments”, supervised by J. Pettré
- Pascal Brault, ENS Rennes - Univ. Rennes 1, “Planning of Minimally Sensitive Trajectories for Mobile Robots”, supervised by Q. Delamare and P. Robuffo Giordano
- Valerio Paduani, Univ. of Rome “La Sapienza”, Italy, “Perception-Aware Human-Assisted Navigation of Mobile Robots on Persistent Trajectories”, supervised by M. Cognetti and P. Robuffo Giordano
- Giulia Matarese, Univ. of Pisa, Italy, “Caring about the human operator: haptic shared control for enhanced user comfort in robotic telemanipulation”, supervised by C. Pacchierotti and P. Robuffo Giordano
- Riccardo Arciulo, Univ. of Rome “La Sapienza”, Italy, “Novel Solution For Human-Robot physical interaction in collaborative tasks”, supervised by C. Pacchierotti and P. Robuffo Giordano
- Ahmed Elsherif, École Centrale de Nantes (EMARO), “Shared control of drones with wearable haptic interfaces”, supervised by C. Pacchierotti and M. Aggravi
- Daniel Lemos Estima, University of Twente, “Integration between a grounded haptic interface and a robot-assisted flexible needle insertion system”, supervised by C. Pacchierotti
- David Steeven Villa Salazar, Univ. Federal do Rio Grande do Sul, “Development of Virtual Reality scenes rendered using ultrasound and vibrotactile haptic interfaces”, supervised by C. Pacchierotti
- Samuel Felton, INSA Rennes - Univ. Rennes 1, “Deep Learning for Visual Servoing”, supervised by E. Marchand
- Gatien Gaumerais, ESIR 2 - Univ. Rennes 1, “Visual-servoing with Parrot Bebop 2 drone”, supervised by F. Spindler
- Quentin Zanini, ENS - Univ. Rennes 1, “Design of a library of 3D shapes for haptic rendering using the ultrahaptics device”, supervised by C. Pacchierotti and M. Marchal (Hybrid group)

9.2.4. Juries

- Paolo Robuffo Giordano: Davide Bicego (Ph.D., member, LAAS, Toulouse)
- Julien Pettré: Jennifer Vandoni (Ph.D., member, Paris Saclay) and Maxime Garcia (Ph.D., reviewer, Grenoble Alpes)
- Eric Marchand: Maxime Rousselot (Ph.D. president, univ. de Rennes 1), Dmitry Kuzovkin (Ph.D. president, univ. de Rennes 1), Gilles Simon (HDR, reviewer, Univ. de Lorraine)
- Marie Babel: Lydia Habib (Ph.D., member, LAMIH, Université Polytechnique Hauts de France)
- François Chaumette: Eulalie Coevet (Ph.D., president, Inria Lille), Mateus Laranjeira (Ph.D., reviewer, Cosmer, Toulon), Brahim Tamadazte (HdR, reviewer, Femto, Besançon), Guillaume Caron (HdR, member, MIS, Amiens)

9.3. 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, 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, the power wheelchair simulator, and augmented reality applications. Some of these demonstrations are available as videos on VispTeam YouTube channel (https://www.youtube.com/user/VispTeam/videos).

9.3.1. Interventions

- François Pasteau participated to the Tech'in Vitré forum with an interview in March 2019.
- Solenne Fortun participated to the Societé de gérontologie de l'ouest et du centre (SGOC) congress with an interview in May 2019.
- Marie Babel animated an interactive session organized in Inria Rennes about "J'peux pas, j'ai informatique" on April 3rd, 2019 for secondary-school pupils.

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] Q. DELAMARE.*Algorithms for estimation and control of quadrotors in physical interaction with their environment*, Univ Rennes, Inria, CNRS, IRISA, France, December 2019, https://tel.archives-ouvertes.fr/tel-02410023

[2] S. JIDDI.*Photometric registration of indoor real scenes using an RGB-D camera with application to mixed reality*, Université Rennes 1, January 2019, https://tel.archives-ouvertes.fr/tel-02167109

[3] I.-F. KENMOGNE.*Set-membership localization of drones using interval methods*, Université de Rennes 1, March 2019, https://tel.archives-ouvertes.fr/tel-02405857

[4] A. LÓPEZ.*Optical flow-based navigation algorithms for virtual humans*, Université de Rennes 1, December 2019, https://tel.archives-ouvertes.fr/tel-02418923

Articles in International Peer-Reviewed Journal

[5] F. ABI-FARRAJ, B. HENZE, C. OTT, P. ROBUFFO GIORDANO, M. A. ROA.*Torque-Based Balancing for a Humanoid Robot Performing High-Force Interaction Tasks*, in "IEEE Robotics and Automation Letters", April 2019, vol. 4, n^{o} 2, p. 2023-2030, (also presented at ICRA'19) [*DOI :* 10.1109/LRA.2019.2898041], https://hal.inria.fr/hal-02051454

[6] F. ABI-FARRAJ, C. PACCHIEROTTI, O. ARENZ, G. NEUMANN, P. ROBUFFO GIORDANO.*A Haptic Shared-Control Architecture for Guided Multi-Target Robotic Grasping*, in "IEEE Transactions on Haptics (ToH)", April 2019, p. 1-16 [*DOI :* 10.1109/TOH.2019.2913643], https://hal.inria.fr/hal-02113206

[7] G. S. AGLIETTI, B. TAYLOR, S. FELLOWES, S. AINLEY, D. TYE, C. COX, A. ZARKESH, A. MAFFICINI, N. VINKOFF, K. BASHFORD, T. SALMON, I. RETAT, C. BURGESS, A. HALL, T. CHABOT, K. KANANI, A. PISSELOUP, C. BERNAL, F. CHAUMETTE, A. POLLINI, W. STEYN.*Survey Paper RemoveDEBRIS: An in-orbit demonstration of technologies for the removal of space debris*, in "Aeronautical Journal -New Series-", 2020, p. 1-23 [*DOI :* 10.1017/AER.2019.136], https://hal.inria.fr/hal-02406580

[8] G. AGLIETTI, B. TAYLOR, S. FELLOWES, T. SALMON, I. RETAT, A. HALL, T. CHABOT, A. PISSELOUP, C. M. COX, A. ZARKESH, A. MAFFICINI, N. VINKOFF, K. BASHFORD, C. BERNAL, F. CHAUMETTE, A. POLLINI, W. STEYN.*The active space debris removal mission RemoveDebris. Part 2: in orbit operations*, in "Acta Astronautica", September 2019, p. 1-16 [*DOI :* 10.1016/J.ACTAASTRO.2019.09.001], https://hal.inria.fr/hal-02286751

[9] S. BRIOT, P. ROBUFFO GIORDANO.*Physical Interpretation of Rigidity for Bearing Formations: Application to Mobility and Singularity Analyses*, in "Journal of Mechanisms and Robotics", April 2019, vol. 11, n^o 3, p. 031006-1–031006-10 [*DOI :* 10.1115/1.4043050], https://hal.archives-ouvertes.fr/hal-02061643

[10] M. CHAUHAN, N. DESHPANDE, C. PACCHIEROTTI, L. MELI, D. PRATTICHIZZO, D. CALDWELL, L. S. MATTOS.*A robotic microsurgical forceps for transoral laser microsurgery*, in "International Journal of Computer Assisted Radiology and Surgery", 2019, vol. 14, n^o 2, p. 321-333 [*DOI :* 10.1007/s11548-018-1887-3], https://hal.inria.fr/hal-01940752

[11] F. CHINELLO, M. MALVEZZI, D. PRATTICHIZZO, C. PACCHIEROTTI.*A modular wearable finger interface for cutaneous and kinesthetic interaction: control and evaluation*, in "IEEE Transactions on Industrial Electronics", January 2020, vol. 67, n^o 1, p. 706-716 [*DOI :* 10.1109/TIE.2019.2899551], https://hal.inria.fr/hal-02021347

[12] N. CROMBEZ, E. M. MOUADDIB, G. CARON, F. CHAUMETTE.*Visual Servoing with Photometric Gaussian Mixtures as Dense Feature*, in "IEEE Transactions on Robotics", February 2019, vol. 35, n^o 1, p. 49-63 [*DOI :* 10.1109/TRO.2018.2876765], https://hal.inria.fr/hal-01896859

[13] L. DEVIGNE, M. AGGRAVI, M. BIVAUD, N. BALIX, S. TEODORESCU, T. CARLSON, T. SPRETERS, C. PACCHIEROTTI, M. BABEL.*Power wheelchair navigation assistance using wearable vibrotactile haptics*, in "IEEE Transactions on Haptics (ToH)", January 2020, p. 1-6 [*DOI :* 10.1109/TOH.2019.2963831], https://hal.inria.fr/hal-02428307

[14] L.-A. DUFLOT, R. REISENHOFER, B. TAMADAZTE, N. ANDREFF, A. KRUPA.*Wavelet and Shearlet-based Image Representations for Visual Servoing*, in "The International Journal of Robotics Research", April 2019, vol. 38, n^o 4, p. 422-450 [*DOI :* 10.1177/0278364918769739], https://hal.archives-ouvertes.fr/hal-01735241

[15] P. ECORMIER-NOCCA, J. PETTRÉ, P. MEMARI, M.-P. CANI.*Image-based Authoring of Herd Animations*, in "Computer Animation and Virtual Worlds", 2019, vol. 30, n^o 3-4, p. 1-11, forthcoming [*DOI :* 10.1002/CAV.1903], https://hal.inria.fr/hal-02127824

[16] J. FORSHAW, G. AGLIETTI, S. FELLOWES, T. SALMON, I. RETAT, A. HALL, T. CHABOT, A. PISSELOUP, D. TYE, C. BERNAL, F. CHAUMETTE, A. POLLINI, W. STEYN.*The active space debris removal mission RemoveDebris. Part 1: from concept to launch*, in "Acta Astronautica", 2019, p. 1-19 [*DOI :* 10.1016/J.ACTAASTRO.2019.09.002], https://hal.inria.fr/hal-02286651

[17] A. FRANCHI, P. ROBUFFO GIORDANO, G. MICHIELETTO.*Online Leader Selection for Collective Tracking and Formation Control: the Second Order Case*, in "IEEE Transactions on Control of Network Systems", December 2019, vol. 6, n^o 4, p. 1415-1425 [*DOI :* 10.1109/TCNS.2019.2891011], https://hal.laas.fr/hal-01964754

[18] C. GAZ, M. COGNETTI, A. OLIVA, P. ROBUFFO GIORDANO, A. DE LUCA.*Dynamic Identification of the Franka Emika Panda Robot With Retrieval of Feasible Parameters Using Penalty-Based Optimization*, in "IEEE Robotics and Automation Letters", October 2019, vol. 4, n^o 4, p. 4147-4154, (also presented at ICRA'19) [*DOI :* 10.1109/LRA.2019.2931248], https://hal.inria.fr/hal-02265293

[19] T. HOWARD, M. MARCHAL, A. LÉCUYER, C. PACCHIEROTTI.*PUMAH : Pan-tilt Ultrasound Mid-Air Haptics for larger interaction workspace in virtual reality*, in "IEEE Transactions on Haptics (ToH)", January 2020, p. 1-6 [*DOI :* 10.1109/TOH.2019.2963028], https://hal.inria.fr/hal-02424247

[20] I.-F. KENMOGNE, V. DREVELLE, E. MARCHAND.*Cooperative Localization of Drones by using Interval Methods*, in "Acta Cybernetica", 2019, p. 1-16, https://hal.inria.fr/hal-02339451

[21] A. KHEDDAR, S. CARON, P. GERGONDET, A. COMPORT, A. TANGUY, C. OTT, B. HENZE, G. MESESAN, J. ENGLSBERGER, M. A. ROA, P.-B. WIEBER, F. CHAUMETTE, F. SPINDLER, G. ORIOLO, L. LANARI, A. ESCANDE, K. CHAPPELLET, F. KANEHIRO, P. RABATE.*Humanoid robots in aircraft manufacturing*, in "IEEE Robotics and Automation Magazine", December 2019, vol. 26, n^{o} 4, p. 30-45 [*DOI :* 10.1109/MRA.2019.2943395], https://hal-lirmm.ccsd.cnrs.fr/lirmm-02303117

[22] A. LÓPEZ, F. CHAUMETTE, E. MARCHAND, J. PETTRÉ.*Character navigation in dynamic environments based on optical flow*, in "Computer Graphics Forum", May 2019, vol. 38, n^{o} 2, p. 181-192, Eurographics 2019 proceedings [*DOI :* 10.1111/CGF.13629], https://hal.inria.fr/hal-02052554

[23] E. MARCHAND.*Subspace-based Direct Visual Servoing*, in "IEEE Robotics and Automation Letters", July 2019, vol. 4, n^{o} 3, p. 2699-2706, (also presented at IROS'19) [*DOI :* 10.1109/LRA.2019.2916263], https://hal.inria.fr/hal-02123993

[24] E. MARCHAND.*Direct visual servoing in the frequency domain*, in "IEEE Robotics and Automation Letters", April 2020, vol. 5, n^{o} 2, p. 620-627 [*DOI :* 10.1109/LRA.2020.2965027], https://hal.inria.fr/hal-02290424

[25] B. PENIN, P. ROBUFFO GIORDANO, F. CHAUMETTE.*Minimum-Time Trajectory Planning Under Intermittent Measurements*, in "IEEE Robotics and Automation Letters", January 2019, vol. 4, n^{o} 1, p. 153-160, (also presented at ICRA'19) [*DOI :* 10.1109/LRA.2018.2883375], https://hal.inria.fr/hal-01935466

[26] D. PRATTICHIZZO, M. OTADUY, H. KAJIMOTO, C. PACCHIEROTTI.*Wearable and Hand-Held Haptics*, in "IEEE Transactions on Haptics (ToH)", July 2019, vol. 12, n^{o} 3, p. 227-231 [*DOI :* 10.1109/TOH.2019.2936736], https://hal.inria.fr/hal-02302243

[27] P. SALARIS, M. COGNETTI, R. SPICA, P. ROBUFFO GIORDANO.*Online Optimal Perception-Aware Trajectory Generation*, in "IEEE Transactions on Robotics", 2019, p. 1-16 [*DOI :* 10.1109/TRO.2019.2931137], https://hal.inria.fr/hal-02278900

[28] S. VILLA SALAZAR, C. PACCHIEROTTI, X. DE TINGUY, A. MACIEL, M. MARCHAL.*Altering the Stiffness, Friction, and Shape Perception of Tangible Objects in Virtual Reality Using Wearable Haptics*, in "IEEE Transactions on Haptics (ToH)", January 2020 [*DOI :* 10.1109/TOH.2020.2967389], https://hal.inria.fr/hal-02450301

[29] E. M. YOUNG, D. GUEORGUIEV, K. J. KUCHENBECKER, C. PACCHIEROTTI.*Compensating for Fingertip Size to Render Tactile Cues More Accurately*, in "IEEE Transactions on Haptics (ToH)", January 2020 [*DOI :* 10.1109/TOH.2020.2966993], https://hal.inria.fr/hal-02443460

[30] Z. ZAKE, F. CHAUMETTE, N. PEDEMONTE, S. CARO.*Vision-Based Control and Stability Analysis of a Cable-Driven Parallel Robot*, in "IEEE Robotics and Automation Letters", February 2019, vol. 4, n^{o} 2, p. 1029-1036, (also presented ar ICRA'19) [*DOI :* 10.1109/LRA.2019.2893611], https://hal.archives-ouvertes.fr/hal-01987856

[31] Z. Zake, F. Chaumette, N. Pedemonte, S. Caro.*Robust 2 1/2D Visual Servoing of a Cable-Driven Parallel Robot Thanks to Trajectory Tracking*, in "IEEE Robotics and Automation Letters", January 2020, vol. 5, n^{o} 2, p. 660-667, https://arxiv.org/abs/2001.06324 [*DOI :* 10.1109/LRA.2020.2965033], https://hal.archives-ouvertes.fr/hal-02429717

[32] K. Zhang, F. Chaumette, J. Chen.*Trifocal tensor-based 6-DOF visual servoing*, in "The International Journal of Robotics Research", September 2019, vol. 38, n^{o} 10-11, p. 1208-1228 [*DOI :* 10.1177/0278364919872544], https://hal.inria.fr/hal-02276270

Invited Conferences

[33] F. Chaumette.*Geometric and end-to-end robot vision-based control*, in "RoMoCo 2019 - 12th Int. Workshop on Robot Motion and Control (Plenary talk)", Poznan, Poland, July 2019, https://hal.inria.fr/hal-02406552

International Conferences with Proceedings

[34] J. Amirian, J.-B. Hayet, J. Pettré.*Social Ways: Learning Multi-Modal Distributions of Pedestrian Trajectories with GANs*, in "CVPR 2019 - IEEE Conference on Computer Vision and Pattern Recognition", Long Beach, United States, IEEE, June 2019, p. 1-9, https://hal.inria.fr/hal-02108756

[35] J. Amirian, W. Van Toll, J.-B. Hayet, J. Pettré.*Data-Driven Crowd Simulation with Generative Adversarial Networks*, in "CASA '19 - International Conference on Computer Animation and Social Agents", Paris, France, March 2019, p. 1-4 [*DOI :* 10.1145/3328756.3328769], https://hal.inria.fr/hal-02134282

[36] F. Berton, A.-H. Olivier, J. Bruneau, L. Hoyet, J. Pettré.*Studying Gaze Behaviour During Collision Avoidance With a Virtual Walker: Influence of the Virtual Reality Setup*, in "VR 2019 - 26th IEEE Conference on Virtual Reality and 3D User Interfaces", Osaka, Japan, IEEE, March 2019, p. 717-725 [*DOI :* 10.1109/VR.2019.8798204], https://hal.inria.fr/hal-02058360

[37] J. Bimbo, C. Pacchierotti, N. G. Tsagarakis, D. Prattichizzo.*Collision detection and isolation on a robot using joint torque sensing*, in "IROS 2019 - IEEE/RSJ International Conference on Intelligent Robots and Systems", Macau, Macau SAR China, IEEE, November 2019, p. 1-6, https://hal.inria.fr/hal-02197009

[38] J. Chevrie, A. Krupa, M. Babel.*Real-time Teleoperation of Flexible Beveled-tip Needle Insertion using Haptic Force Feedback and 3D Ultrasound Guidance*, in "ICRA 2019 - IEEE International Conference on Robotics and Automation", Montreal, Canada, IEEE, May 2019, p. 1-7, https://hal.inria.fr/hal-02053101

[39] J. G. Da Silva Filho, A.-H. Olivier, A. Crétual, J. Pettré, T. Fraichard.*Effective Human-Robot Collaboration in near symmetry collision scenarios*, in "RO-MAN 2019 - 28th IEEE International Conference on Robot & Human Interactive Communication", New Dehli, India, IEEE, October 2019, p. 1-8, https://hal.inria.fr/hal-02267705

[40] X. De Tinguy, C. Pacchierotti, M. Emily, M. Chevalier, A. Guignardat, M. Guillaudeux, C. Six, A. Lécuyer, M. Marchal.*How different tangible and virtual objects can be while still feeling the same?*, in "WHC 2019 - IEEE World Haptics Conference", Tokyo, Japan, IEEE, July 2019, p. 1-6 [*DOI :* 10.1109/WHC.2019.8816164], https://hal.inria.fr/hal-02121839

[41] X. De Tinguy, C. Pacchierotti, M. Marchal, A. Lécuyer.*Toward Universal Tangible Objects: Optimizing Haptic Pinching Sensations in 3D Interaction*, in "VR 2019 - 26th IEEE Conference on Virtual Re-

ality and 3D User Interfaces", Osaka, Japan, IEEE, 2019, p. 321-330 [*DOI :* 10.1109/VR.2019.8798205], https://hal.inria.fr/hal-02021319

[42] L. DEVIGNE, F. PASTEAU, T. CARLSON, M. BABEL.*A shared control solution for safe assisted power wheelchair navigation in an environment consisting of negative obstacles: a proof of concept*, in "SMC 2019 - IEEE International Conference on Systems, Man, and Cybernetics", Bari, Italy, IEEE, October 2019, p. 1043-1048 [*DOI :* 10.1109/SMC.2019.8914211], https://hal.inria.fr/hal-02263532

[43] H. GURNEL, M. MARCHAL, L. LAUNAY, L. BEUZIT, A. KRUPA.*Design of haptic guides for pre-positioning assistance of a comanipulated needle*, in "SMC 2019 - IEEE International Conference on Systems, Man, and Cybernetics", Bari, Italy, IEEE, October 2019, p. 478-485 [*DOI :* 10.1109/SMC.2019.8914395], https://hal.inria.fr/hal-02387192

[44] H. GURNEL, M. MARCHAL, L. LAUNAY, L. BEUZIT, A. KRUPA.*Preliminary evaluation of haptic guidance for pre-positioning a comanipulated needle*, in "SURGETICA 2019", Rennes, France, June 2019, p. 1-3, https://hal.inria.fr/hal-02387214

[45] T. HOWARD, G. GALLAGHER, A. LÉCUYER, C. PACCHIEROTTI, M. MARCHAL.*Investigating the recognition of local shapes using mid-air ultrasound haptics*, in "WHC 2019 - IEEE World Haptics Conference", Tokyo, Japan, IEEE, July 2019, p. 1-6 [*DOI :* 10.1109/WHC.2019.8816127], https://hal.inria.fr/hal-02121329

[46] A. LÓPEZ, F. CHAUMETTE, E. MARCHAND, J. PETTRÉ.*Attracted by light: vision-based steering virtual characters among dark and light obstacles*, in "MIG 2019 - ACM SIGGRAPH Conference Motion Interaction and Games", Newcastle upon Tyne, United Kingdom, ACM, October 2019, p. 1-6 [*DOI :* 10.1145/3359566.3360085], https://hal.inria.fr/hal-02299397

[47] E. MARCHAND, F. CHAUMETTE, T. CHABOT, K. KANANI, A. POLLINI.*RemoveDebris Vision-Based Navigation preliminary results*, in "IAC 2019 - 70th International Astronautical Congress", Washington D.C., United States, October 2019, p. 1-10, https://hal.inria.fr/hal-02315122

[48] B. NIAY, A.-H. OLIVIER, J. PETTRÉ, L. HOYET.*The Influence of Step Length to Step Frequency Ratio on the Perception of Virtual Walking Motions*, in "MIG 2019 - 12th annual ACM SIGGRAPH conference on Motion, Interaction and Games", Newcastle upon Tyne, United Kingdom, ACM Press, October 2019, p. 1-2 [*DOI :* 10.1145/3359566.3364687], https://hal.archives-ouvertes.fr/hal-02378300

[49] A.-H. OLIVIER, N. LE BORGNE, M. BABEL, A. CRÉTUAL, J. PETTRÉ.*Evitement de collision entre un piéton et une personne sur un fauteuil roulant motorisé*, in "SOFPEL 2019 - 26ème congrès de la Société Francophone Posture, Équilibre et Locomotion", Montréal, Canada, December 2019, https://hal.inria.fr/hal-02373525

[50] R. RAHAL, F. ABI-FARRAJ, P. R. GIORDANO, C. PACCHIEROTTI.*Haptic Shared-Control Methods for Robotic Cutting under Nonholonomic Constraints*, in "IROS 2019 - IEEE/RSJ International Conference on Intelligent Robots and Systems", Macau, Macau SAR China, IEEE, November 2019, p. 1-7, https://hal.inria.fr/hal-02197603

[51] M. SELVAGGIO, P. ROBUFFO GIORDANO, F. FICUCIELLO, B. SICILIANO.*Passive Task-Prioritized Shared-Control Teleoperation with Haptic Guidance*, in "ICRA 2019 - IEEE International Conference on Robotics and Automation", Montreal, Canada, IEEE, May 2019, p. 1-7, https://hal.inria.fr/hal-02051476

[52] A. SENGUPTA, A. KRUPA, E. MARCHAND.*RGB-D tracking of complex shapes using coarse object models*, in "ICIP 2019 - IEEE International Conference on Image Processing", Taipei, Taiwan, IEEE, September 2019, p. 1-5 [*DOI :* 10.1109/ICIP.2019.8803574], https://hal.inria.fr/hal-02129243

[53] A. SENGUPTA, A. KRUPA, E. MARCHAND.*Tracking of Non-Rigid Objects using RGB-D Camera*, in "SMC 2019 - IEEE International Conference on Systems, Man, and Cybernetics", Bari, Italy, IEEE, October 2019, p. 3310-3317 [*DOI :* 10.1109/SMC.2019.8914543], https://hal.inria.fr/hal-02178353

[54] G. VAILLAND, F. GRZESKOWIAK, L. DEVIGNE, Y. GAFFARY, B. FRAUDET, E. LEBLONG, F. NOUVIALE, F. PASTEAU, R. LE BRETON, S. GUEGAN, V. GOURANTON, B. ARNALDI, M. BABEL.*User-centered design of a multisensory power wheelchair simulator: towards training and rehabilitation applications*, in "ICORR 2019 - International Conference on Rehabilitation Robotics", Toronto, Canada, 2019, p. 1-6, https://hal.inria.fr/hal-02134530

[55] W. VAN TOLL, J. PETTRÉ.*Connecting Global and Local Agent Navigation via Topology*, in "MIG 2019 - ACM SIGGRAPH Conference Motion Interaction and Games", Newcastle upon Tyne, United Kingdom, ACM, October 2019, p. 1-10 [*DOI :* 10.1145/3359566.3360084], https://hal.inria.fr/hal-02308297

[56] Z. ZAKE, S. CARO, A. SUAREZ ROOS, F. CHAUMETTE, N. PEDEMONTE.*Stability Analysis of Pose-Based Visual Servoing Control of Cable-Driven Parallel Robots*, in "CableCon 2019 - Fourth International Conference on Cable-Driven Parallel Robots", Krakow, Poland, June 2019, p. 1-12 [*DOI :* 10.1007/978-3-030-20751-9_7], https://hal.archives-ouvertes.fr/hal-02127451

National Conferences with Proceeding

[57] T. DUVERNE, T. ROUGNANT, F. LE YONDRE, F. BERTON, J. BRUNEAU, L. HOYET, J. PETTRÉ, A.-H. OLIVIER.*Analyse des réactions corporelles à la transgression des normes de proxémie en contexte de spectacle sportif : expérimentations en situations réelles et virtuelles*, in "ACAPS 2019 - 18ème congrès de l'Association des Chercheurs en Activités Physiques et Sportives", Paris, France, October 2019, https://hal.inria.fr/hal-02393360

Conferences without Proceedings

[58] L. DEVIGNE, M. BABEL, R. LE BRETON, E. LEBLONG, V. GOURANTON, F. PASTEAU, B. FRAUDET, S. GUEGAN, G. VAILLAND, Y. GAFFARY.*Expérience de conduite en fauteuil roulant dans une ville virtuelle avec un simulateur multisensoriel conçu selon une approche centrée sur l'utilisateur : une étude de cas*, in "SOFMER 2019 - 34ème congrès de la Société Français de Médecine Physique et de Réadaptation", Bordeaux, France, October 2019, https://hal.inria.fr/hal-02339573

[59] P. GALIEN, S. ACHILLE-FAUVEAU, E. LEBLONG, A. DURUFLE, M. BABEL, B. FRAUDET, N. BENOIT.*Vision croisée sur la robotique d'assistance dans la vie quotidienne des patients et des aidants*, in "SOFMER 2019 - 34ème congrès de la Société Français de Médecine Physique et de Réadaptation", Bordeaux, France, October 2019, https://hal.inria.fr/hal-02339550

[60] S. LYNCH, R. KULPA, L. A. MEERHOFF, A. SOREL, J. PETTRÉ, A.-H. OLIVIER.*Collision avoidance between walkers with a twist: strategies for curvilinear and rectilinear paths*, in "ISPGR 2019 - Conference of the International Society for Posture & Gait Research", Edinburgh, United Kingdom, June 2019, https://hal.inria.fr/hal-02058340

[61] B. NIAY, A.-H. OLIVIER, J. PETTRÉ, L. HOYET.*The Influence of Step Length to Step Frequency Ratio on the Perception of Virtual Walking Motions*, in "ACM SIGGRAPH Symposium on Applied Perception", Barcelone, Spain, September 2019, https://hal.archives-ouvertes.fr/hal-02395303

[62] A.-H. OLIVIER, N. LE BORGNE, M. BABEL, A. CRÉTUAL, J. PETTRÉ.*Collision avoidance between a walker and a person on an electric powered wheelchair*, in "ISPGR 2019 - Conference of the International Society for Posture & Gait Research", Edinburgh, United Kingdom, June 2019, https://hal.inria.fr/hal-02058342

Scientific Books (or Scientific Book chapters)

[63] F. CHAUMETTE.*Robot Visual Control*, in "Encyclopedia of Systems and Control, 2nd edition", J. BAILLIEUL, T. SAMAD (editors), Springer London, December 2020, p. 1-18 [*DOI :* 10.1007/978-1-4471-5102-9_170-2], https://hal.inria.fr/hal-02422019

[64] E. MARCHAND.*Visual Tracking*, in "Encyclopedia of Robotics", Springer, 2020, p. 1-16, forthcoming, https://hal.inria.fr/hal-02426694

[65] A. SPADA, M. COGNETTI, A. DE LUCA.*Locomotion and Telepresence in Virtual and Real Worlds*, in "Locomotion and Telepresence in Virtual and Real Worlds", S. P. IN ADVANCED ROBOTICS (editor), June 2019, p. 85-98 [*DOI :* 10.1007/978-3-319-89327-3_7], https://hal.inria.fr/hal-02265284

Other Publications

[66] C. GAZ, M. COGNETTI, A. OLIVA, P. ROBUFFO GIORDANO, A. DE LUCA.*Dynamic Identification of the Franka Emika Panda Robot with Retrieval of Feasible Parameters Using Penalty-Based Optimization*, October 2019, p. 1-9, Supplementary material, https://hal.inria.fr/hal-02265294

Project-Team SERPICO

Space-timE RePresentation, Imaging and cellular dynamics of molecular COmplexes

IN COLLABORATION WITH: UMR 144 - Compartimentation et dynamique cellulaires

IN PARTNERSHIP WITH:
CNRS

Institut Curie

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Computational Biology

Table of contents

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
A3.3. - Data and knowledge analysis
A3.3.3. - Big data analysis
A3.4. - Machine learning and statistics
A3.4.1. - Supervised learning
A3.4.5. - Bayesian methods
A3.4.6. - Neural networks
A3.4.7. - Kernel methods
A3.4.8. - Deep learning
A5.3. - Image processing and analysis
A5.3.2. - Sparse modeling and image representation
A5.3.3. - Pattern recognition
A5.3.4. - Registration
A5.4.1. - Object recognition
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
A5.9.5. - Sparsity-aware processing
A5.9.6. - Optimization tools
A6.1.2. - Stochastic Modeling
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. - 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
A9.2. - Machine learning
A9.3. - Signal analysis

Other Research Topics and Application Domains:

B1.1.1. - Structural biology
B1.1.7. - Bioinformatics
B1.1.8. - Mathematical biology
B2.2.3. - Cancer
B2.6. - Biological and medical imaging

1. Team, Visitors, External Collaborators

Research Scientists

Charles Kervrann [Team leader, Inria, Senior Researcher, HDR]
Jean Salamero [Team leader, CNRS, Senior Researcher, HDR]
Patrick Bouthemy [Inria, Senior Researcher, HDR]

External Collaborators

Frédéric Lavancier [External collaborator, Univ de Nantes, HDR]
Sylvain Prima [Invited researcher, Inria, untile Dec 2019]

Technical Staff

Ludovic Leconte [CNRS, Engineer, from Jan 2019]
Emmanuel Moebel [Inria, Engineer, from Dec 2019]
Sylvain Prigent [CNRS, Engineer, from Jan 2019 until Dec 2019]

PhD Students

Anca-Georgiana Caranfil [Univ de Rennes 1, until Nov 2019]
Gwendal Fouché [Inria, from Oct 2019 (with EPC HYBRID, Inria Rennes)]
Yunjiao Lu [INRA]
Léo Maczyta [Inria]
Sandeep Manandhar [Inria, until Nov 2019]
Emmanuel Moebel [Inria, until Jan 2019]
Antoine Salomon [Inria]

Post-Doctoral Fellows

Anais Badoual [Inria, Post-Doctoral Fellow, from Oct 2019]
Cesar Augusto Valades Cruz [Inria, Post-Doctoral Fellow, from Apr 2019 until Aug 2019, Starting Research Position (IPL Naviscope), from Sep 2019]

Administrative Assistants

Huguette Bechu [Inria, Administrative Assistant]
Michèle Colineau [CNRS - Institut Curie, Administrative Assistant]

2. Overall Objectives

2.1. Glossary

LLSM (Lattice Light Sheet Microscopy): high resolution Structured Illumination Microscopy that uses Light Sheet Bessel Beam illumination [47].

TIRF (Total Internal Reflectance): 2D optical microscopy using evanescent waves and total reflectance [45].

STORM (Stochastic Optical Reconstruction Microscopy): high-resolution microscopy using stochastic photo-activation of fluorophores and adjustment of point spread functions.

PALM (Photo-Activated Localization Microscopy): high-resolution microscopy using stochastic photo-activation of fluorophores and adjustment of point spread functions [46].

Cryo-ET (Cryo-Electron Tomography): 3D representation of sub-cellular and molecular objects of 5-20 nanometers, 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 tens nanometers, in live. As a result, fluorescent microscopy and multimodal imaging (fluorophores at various wavelengths) have become the workhorse of modern biology. As a matter of fact, taking into account all the publications in the 10 most relevant journals in fundamental biology (those with highest IF) for the last 2018-2019 years, the ratio of experimental figures based on BioImage data is close to 70% (GBI EoE.V, Singapour, Sep 2019). 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 long-term 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, modeling 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 CNRS-UMR 144 Institut Curie (and in cooperation with the "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 CNRS-UMR 144 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. 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. 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 analyzing 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 co-orientation 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

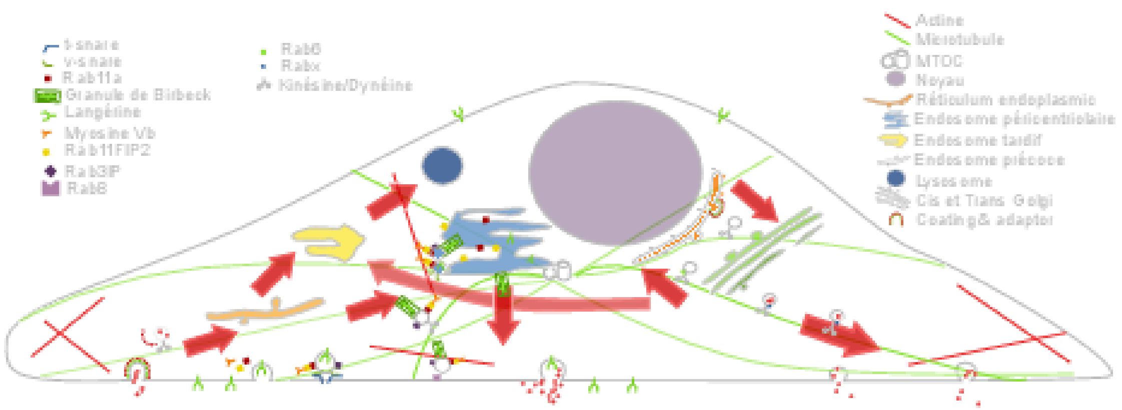


Figure 1. Cargo Langerin Trafficking controlled by Rab11A/Rab11FIP2/MyoVb platform.

In the past recent years, research carried out out together with the “Space Time imaging of Endomembranes and organelles Dynamics” team at CNRS-UMR 144Institut Curie contributed to a better understanding of the intracellular compartmentation, particularly in specialized model cells such as melanocytes and Langerhans cells of the epidermis, of the components and structural events involved in the biogenesis of their specialized organelles: melanosomes and Birbeck granules, respectively and to the understanding on how the dynamics of those structures relate to their physiological functions. These studies have started to highlight: i/ the measurement of multiple sorting and structural events involved in the biogenesis of these organelles; 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, an issue that become more and more accessible, thanks to improvement in all domains related to live imaging. 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, how different machineries evolve together and thus coordinate (e.g. see Fig. 1). They help to decipher cell organization and function at different scales through an integrative workflow of methodological and technological developments. New approaches, such as optogenetics may even help controlling cell functions.

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 differentiation and cell transformation. Our methodological goal is also to link dynamics information obtained through diffraction limited light microscopy, at a time regime compatible with live cell imaging and close to biochemical molecular interactions. The overview of ultrastructural organization will be achieved by complementary electron microscopy methods which have also undergone a revolutionary improvement over the last decade. Image visualization and quantitative analysis are of course essential issues in this context.

4.2. Imaging and analysis of cytoskeleton 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 from primary tumor in blood (extravasation) and eventually the colonization of other organs and the formation of secondary tumors.

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. Another issue is of course to measure and understand how membrane protrusion and retraction keep the cell in homeostasis, which of course relate to membrane traffic.

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 how molecular trafficking is coordinated with cytoskeleton changes in these fundamental cellular functions.

5. Highlights of the Year

5.1. Highlights of the Year

- The SERPICO team organized the 7th International Conference on “Quantitative BioImaging” (QBI, https://www.quantitativebioimaging.com/qbi2019/) in January 2019 (350 attendees) in Rennes. The Quantitative BioImaging conference encourages scientific communication between researchers with interest in quantitative imaging in biological and biomedical sciences. A particular emphasis is to promote interdisciplinary interactions between physicists, computer scientists, chemists, mathematicians, and biologists.
- Emmanuel Moebel and Sandeep Manandhar defended their PhD theses in 2019.
- The DeepFinder algorithm was ranked first at the international SHREC’19 Challenge: "classification in cryo-electron tomograms" (Eurographics Workshop on 3D Object Retrieval – SHREC - 3D Shape Retrieval Contest (2019), Genova, Italy).

6. New Software and Platforms

6.1. GcoPS

KEYWORDS: Photonic imaging - Fluorescence microscopy - Image processing - Statistic analysis

FUNCTIONAL DESCRIPTION: The GcoPS (Geo-Co-Positioning System) software is dedicated to the co-localization 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. Colocalization amounts here 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: Thierry Pécot, Frédéric Lavancier, Charles Kervrann and Liu Zengzhen
- Partners: Université de Nantes - UMR 144 CNRS - Institut Curie - Hollings Cancer Center at the Medical University of South Carolina, Charleston SC, USA
- Contact: Charles Kervrann
- Publication: Testing independence between two random sets for the analysis of colocalization in bioimaging
- URL: http://icy.bioimageanalysis.org/plugin/GcoPS

6.2. THOTH

Testing HypOtheses for diffusion TricHotomy

KEYWORDS: Photonic imaging - Fluorescence microscopy - Biomedical imaging - Classification - Statistical categorisation techniques - Statistics - Image sequence - Visual tracking

FUNCTIONAL DESCRIPTION: The THOTH software classifies biomolecule trajectories of biomolecules (computed with tracking algorithms) into three groups of diffusion: (i) free diffusion, (ii) subdiffusion or (iii) superdiffusion. Brownian motion corresponds to the NULL hypothesis. THOTH is a nonparametric three-decision test whose alternatives are subdiffusion and superdiffusion. Single and multiple testing procedures control respectively the type I error and the false discovery rate. THOTH 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, Charles Kervrann and Myriam Vimond
- Partner: ENSAI
- Contact: Charles Kervrann
- Publication: Statistical analysis of particle trajectories in living cells
- URL: https://team.inria.fr/serpico/software/thot/

6.3. SparseVolution

Sparse Variation for 2D Image Decovolution

KEYWORDS: Fluorescence microscopy - Image processing - Deconvolution - Inverse problem

FUNCTIONAL DESCRIPTION: In order 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 the first (Total Variation) or second (Hessian Variation) derivatives to favor 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 algorithm has been dedicated to deconvolve very large 2D (e.g. 20 000 x 20 000) images or 3D images.

- Participants: Hoai Nam Nguyen, Charles Kervrann, Sylvain Prigent and Cesar Augusto Valades Cruz
- Partners: Innopsys - UMR 144 CNRS - Institut Curie
- Contact: Charles Kervrann

6.4. AiryscanJ

Reconstruction of AiryScan microscopy images

KEYWORDS: Image reconstruction - Fluorescence microscopy - Photonic imaging - Deconvolution - Image analysis

FUNCTIONAL DESCRIPTION: The AiryscanJ software enables to reconstruct a high resolution image from an array of multiple raw images acquired with the Airyscan technology (32 detectors). Airyscanning is a recent technique based on confocal laser scanning microscopy. The AiryscanJ software gathers four reconstruction methods (ISM, IFED, ISFED, ISM-Deconvolution) to compute a high resolution image: 1/ ISM amounts to summing the preliminarily registered raw images. 2/ IFED is the weighted difference between the inner detectors and the outer detector. 3/ ISFED is the weighted difference between the registered outer detectors and the original outer detectors. AiryscanJ automatically estimates the parameter controlling the IFED and ISFED algorithms. 4/ ISM-Deconvolution allows reconstructing a high resolution image by applying a deconvolution algorithm on the ISM image.

- Participants: Sylvain Prigent, Stephanie Dutertre and Charles Kervrann
- Partners: Université de Rennes 1 - CNRS
- Contact: Charles Kervrann
- URL: https://gitlab.inria.fr/serpico/airyscanj_bin

6.5. DeepFinder

Deep learning for macromolecule identification within 3D cellular cryo-electron tomograms

KEYWORDS: Image analysis - Deep learning - Cryo-electron microscopy - Object detection

FUNCTIONAL DESCRIPTION: DeepFinder is a computational approach that uses artificial neural networks to accurately and jointly localize multiple types and/or states of macromolecules in 3D cellular cryo-electron tomograms. DeepFinder leverages deep learning and outperforms the commonly-used template matching method on ideal data. On synthetic image data (SHREC 2019 challenge), DeepFinder is very fast and produces superior detection results when compared to other competitive deep learning methods, especially on small macromolecules. On experimental cryo-ET data depicting ribosomes, the detection results obtained by DeepFinder are consistent with expert annotations. We have got a high overlap of 86% and a similar structure resolution determined by subtomogram averaging.

- Participants: Emmanuel Moebel, Antonio Martinez and Charles Kervrann
- Partners: Max Planck Institute Martinsried - Fondation Fourmentin-Guilbert
- Contact: Charles Kervrann
- URL: https://gitlab.inria.fr/serpico/deep-finder

6.6. CPAnalysis

Change point detection algorithm for detecting switches of diffusion along 2D-3D single particle trajectories

KEYWORDS: Statistical modeling - Statistical physics - Fluorescence microscopy - Motion analysis

FUNCTIONAL DESCRIPTION: The change point detection algorithm is designed for detecting switches of diffusion along a given 2D-3D particle trajectory. We consider that the particle can switch between three main motion modes: 1/ Superdiffusion which occurs when the particle is transported via molecular motors along the cytoskeleton, 2/ Free diffusion (or Brownian motion) which arises when the particle evolves freely inside the cytosol, 3/ Subdiffusion observed when the particle is confined in a domain or evolves in a crowded area. The algorithm is a sequential nonparametric procedure based on a nonparametric three-decision test computed on local sliding windows along the trajectory.

- Participants: Antoine Salomon, Vincent Briane, Myriam Vimond, Jean Salamero and Charles Kervrann
- Partners: ENSAI - UMR 144 CNRS - Institut Curie
- Contact: Charles Kervrann
- Publication: A sequential algorithm to detect diffusion switching along intracellular particle trajectories
- URL: https://team.inria.fr/serpico/software/cpanalysis/

6.7. FlowScope

Optical flow computation for 3D fluorescence microscopy

KEYWORDS: Motion analysis - Fluorescence microscopy - Image analysis - Data visualization

FUNCTIONAL DESCRIPTION: The FLOWSCOPE software is able to estimate 3D motion between two fluorescence microscopy volumes. The underlying variational method amounts to minimizing an energy functional made up of two terms: a data term and a regularization term. The data term is derived from the continuous form of Census signature and the smoothness of the flow field is imposed by a L2 regularization term. The method is implemented for a single core CPU. FLOWSCOPE outputs three separate files corresponding to the motion vector components. The flow fields can be visualized with an appropriate color code named 3PHS. The 3PHS map projects the flow field onto the three orthogonal planes selected by the user. The projections of the vector field is color coded in the Hue (direction) and the Saturation (amplitude) spaces.

- Participants: Patrick Bouthemy, Charles Kervrann, Sylvain Prigent, Leo Maury and Philippe Roudot
- Partner: Danuser lab, University of Texas Southwestern, Dallas, USA
- Contact: Sandeep Manandhar
- Publication: 3D Flow Field Estimation and Assessment for Live Cell Fluorescence Microscopy
- URL: https://gitlab.inria.fr/smanandh/flowscope

6.8. MWR

Missing Wedge Restoration (MWR) and Noise Removal in 3D Cryo-Tomography

KEYWORDS: Image analysis - Inverse problem - Cryo-electron microscopy - Monte-Carlo methods - Bayesian estimation

FUNCTIONAL DESCRIPTION: The Missing Wedge Restoration (MWR) software enables to reduce the high amount of noise and artifacts observed in 3D cellular cryo-tomograms, induced by the presence of a missing wedge (MW) in the spectral domain. MWR takes as input a 3D tomogram derived from limited-angle tomography, and gives as output a 3D denoised and artifact compensated volume. The artifact compensation is achieved by filling up the MW with meaningful information. A Minimum Mean Square Error (MMSE) estimator is computed by applying a dedicated Markov Chain Monte-Carlo (MCMC) sampling procedure based on the Metropolis-Hasting algorithm. MWR can be used to enhance visualization or as a pre-processing step for image analysis, including segmentation and classification of macromolecules.

- Participants: Emmanuel Moebel and Antonio Martinez
- Partners: Max Planck Institute Martinsried - Fondation Fourmentin-Guilbert
- Contact: Charles Kervrann
- Publication: A Monte Carlo framework for missing wedge restoration and noise removal in cryo-electron tomography
- URL: https://gitlab.inria.fr/serpico/mwr

6.9. Platforms

6.9.1. Mobyle@serpico platform and software distribution

Participants: Sylvain Prigent, Léo Maury, Charles Kervrann.

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/). A few of them are freely available on the team website https://team.inria.fr/serpico/software/ under a proprietary license.

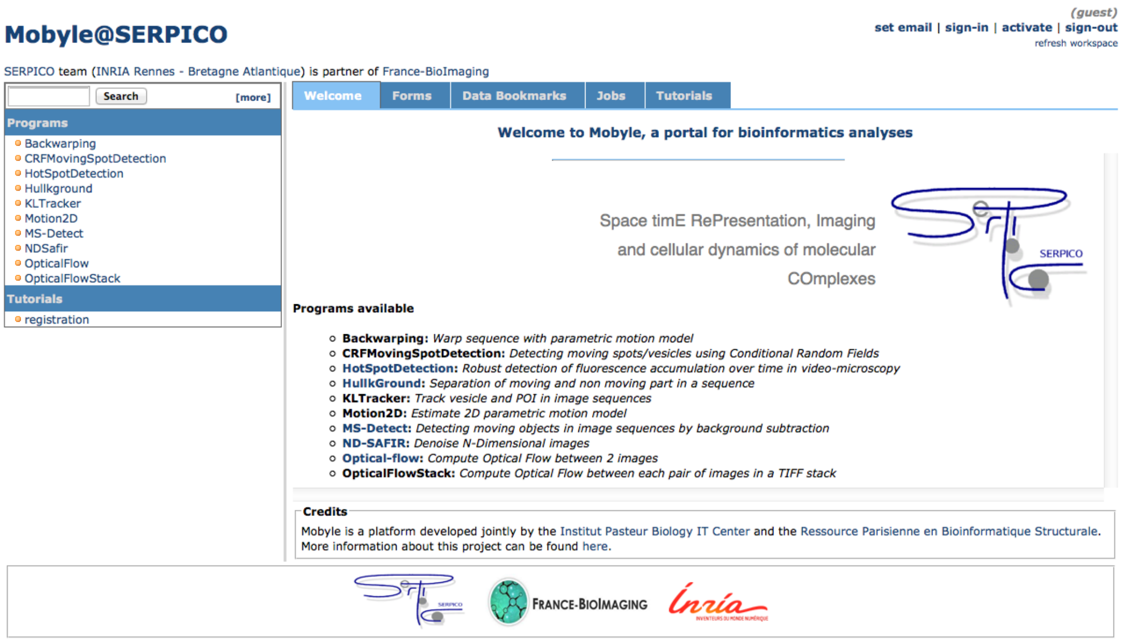


Figure 2. Mobyle@serpico web portal.

Mobyle@serpico web portal: An on-line version (http://mobyle-serpico.rennes.inria.fr) of the image processing algorithms has been developed 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[9], Atlas[1], Hullkground[48], KLTracker[52], Motion2D[51], MS-detect[49], ND-Safir[4], OpticalFlow and Flux Estimation [9]. The web interface makes our image processing methods available for biologists at Mobyle@serpico 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 [4], Hullkground [48], Motion2D [51], Atlas [1]. The C-Craft [9], QuantEv and GcoPS [19] algorithms have been developed for the image processing ICY platform (http://icy.bioimageanalysis.org/).

Partners: CNRS-UMR 144 Institut Curie & France-BioImaging (UMS 3714 CEMIBIO).

6.9.2. Bioimage-IT for bioimage management and processing

Participants: Sylvain Prigent, Cesar Augusto Valades Cruz, Léo Maury, Jean Salamero, Charles Kervrann.

New image acquisition systems generate large number of images and large volume images. Such data sets are hard to store, to process and to analyze for one user in a workstation. Many solutions exist for data management (e.g. Omero, OpenImadis), image analysis (e.g. Fiji, Icy, CellProfiler) and statistics (e.g R). Each of them has its specificities and several bridges have been developed between pieces of software. Nevertheless, in many use-cases, we need to perform analysis using tools that are available in different pieces of software and different languages. It is then tedious to create a workflow that brings the data from one tool to another. It needs programing skills and most of the time, a dedicated script using a dedicated file system for processed data management is developed. The aim of BioImage-IT is to create a "bandmaster" application that allow any scientist to annotate, process, and analyze data using only one single high level application. This BioImage-IT application is based on 3 components:

- an image annotation method based on a JSON file system,
- an image processing and analysis tools integration method based on Docker and XML commands description,
- an application with a graphical interface to easily annotate data, run processing tools, and visualize data and results.

This software architecture has three main goals. First, data are annotated using a file system. This means that data are not dependent on any software like a SQL database, and each experiment can then be stored in a different directory and can be moved from one server to another or to any drive with a simple copy pasting operation. Second, the processing tools are used as binary packages managed by the Docker technology. Docker enables to gently handle dependencies and several versions of the same tool. Any existing tool can then be integrated in its native programming language. Third, using a single "bandmaster" application allows one to automatically generate metadata for any processed data, improving the traceability and the repeatability of any experimental result.

BioImage-IT (https://project.inria.fr/bioimageit) is developed in the context of the France-BioImaging research infrastructure in coordination with the IPDM-FBI (Image Processing and Data Management) node in order to provide a standardized image processing tool set and data management for the imaging facilities.

Partners: CNRS-UMR 144 and U1143 INSERM/CNRS-UMR 3666, Institut Curie & France-BioImaging (UMS 3714 CEMIBIO).

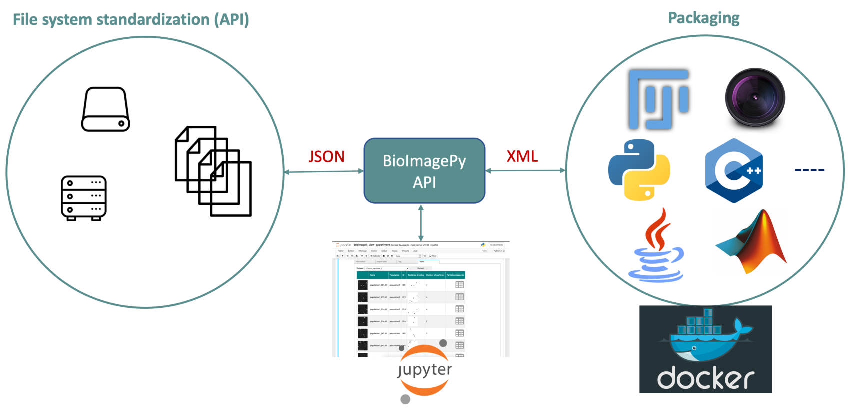


Figure 3. Scheme of the BioImage-IT components interactions.

7. New Results

7.1. Empirical SURE-guided microscopy super-resolution image reconstruction from confocal multi-array detectors

Participants: Sylvain Prigent, Charles Kervrann.

Recent confocal microscopes use an array detector instead of single point detector to take multiple views of the same sample. The microscope output is then an array of images, one image per detector. The array of images is then processed to build a single image with higher signal-to-noise ratio and higher resolution than a classical confocal microscope image. In the literature, several methods have been recently proposed to reconstruct the single high resolution image: i/ the ISM method combines array registration and Wiener deconvolution; ii/ the IFED method estimates a high resolution image by subtracting the outer detectors of the array to the inner detectors; iii/ the ISFED consists in subtracting the outer registered detectors and the outer raw images. In that context, we proposed a SURE-guided (Stein's unbiased risk estimation) estimation method to automatically select the parameter ϵ controlling the IFED and ISFED reconstruction algorithms (see Figure 4). We showed on real data that the proposed method is capable to achieve a resolution close to 100 nm without any deconvolution method.

Software: AiryscanJ (see Section 6.4).
Collaborator: S. Dutertre (IGDR – Institute Genetics & Development of Rennes).

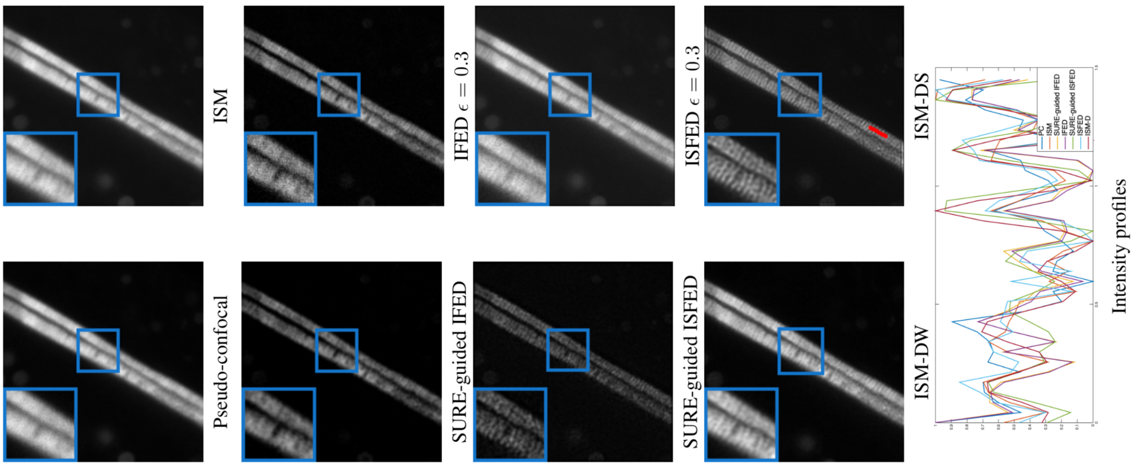


Figure 4. Results obtained on a c.elegans sample and intensity profiles for all the tested methods. The profile line is shown with a red line on the ISM-DS image (top right).

7.2. Dense mapping of intracellular diffusion and drift from single-particle tracking data analysis

Participants: Antoine Salomon, Cesar Augusto Valades Cruz, Charles Kervrann.

It is of primary interest for biologists to be able to locally estimate diffusion and drift inside a cell. In our framework, we assumed that particle motion is governed by the following Langevin equation: $\mathbf{dx} = \mathbf{b}(\mathbf{x})dt + \sigma(\mathbf{x})\mathbf{dw}$ where $\mathbf{x}$ denotes the 2D or 3D spatial coordinates of the particle, $\mathbf{b}$ the drift vector,

σ the diffusion coefficient, and $\mathbf{w}$ the standard Gaussian white noise. In that context, we proposed a new mapping method inspired from [50] that developed a method providing results in the form of matrices by scanning the data by blocks, and from the framework in [5], dedicated to both classifying particle motion types and detecting potential motion switches along a trajectory. To avoid the calculation of both drift and diffusion in cell coordinates where no data is available, we replaced the scanning movement of an averaging window by a Gaussian window centered on trajectory points. Each drift vector and each diffusion coefficient are calculated at coordinates corresponding exactly to the coordinates given by the preliminary particle tracking, which provides more details. A nonparametric three-decision test enables to label trajectories or sub-trajectories [5]. This information is then used. to calculate drift and diffusion coefficient (or Kramers-Moyal coefficients) maps separately on each class of motion with the most appropriate diffusion models: confined motion (the particle is bound to a specific point), Brownian motion (the particle moves randomly), and directed motion (the particle moves in a given direction) (see Figure 5).

Software: THOTH and CPAnalysis (see Sections 6.2 and 6.6).

Collaborators: V. Briane (UNSW Sydney, School of Medical Sciences, Australia),
L. Leconte and J. Salamero (CNRS-UMR 144, Institut Curie, PSL Research University),
L. Johannes (U1143 INSERM / CNRS-UMR 3666, Institut Curie, PSL Research University),
E. Derivery (MRC laboratory of Molecular Biology, Cambridge, UK),
L. Muresan (Cambridge Advanced Imaging Centre, UK).

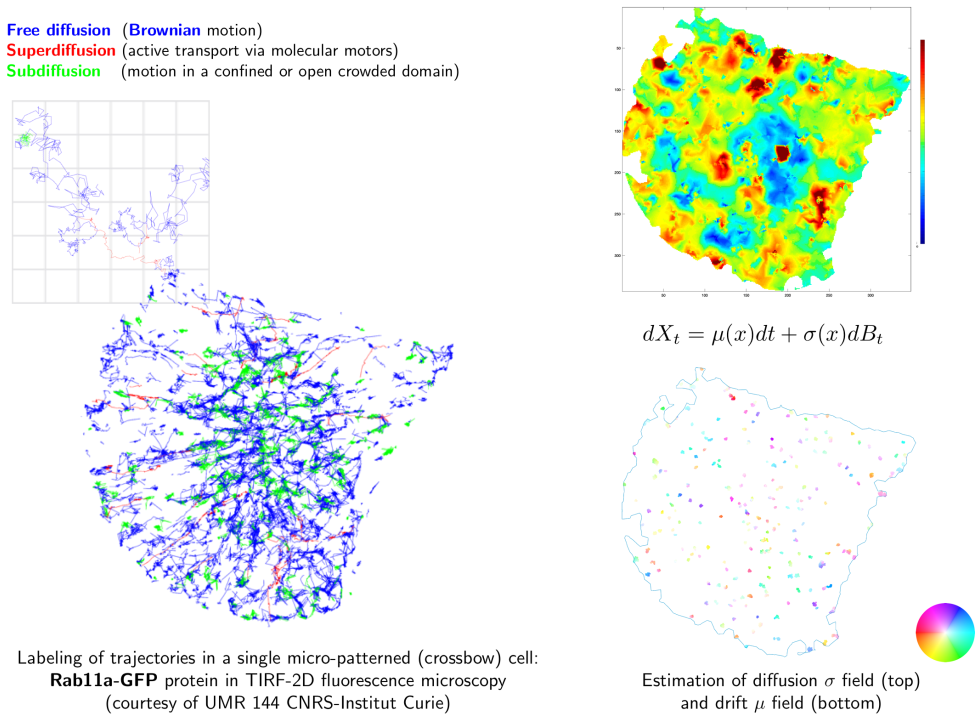


Figure 5. Intracellular diffusion and drift maps estimated from simulated tracking data (FluoSIM).

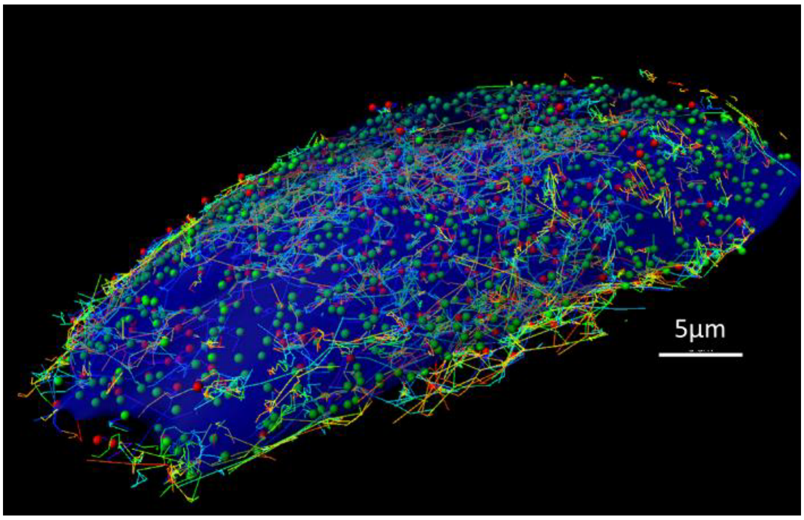


Figure 6. 3D tracking of Gal3-Atto647n (red) vs AP2-eGFP (green) adaptor protein in SUM159 cell.

7.3. 3D tracking of endocytic and exocytic events using lattice light sheet microscopy

Participants: Cesar Augusto Valades Cruz, Ludovic Leconte, Jean Salamero, Charles Kervrann.

The study of the whole cell dynamics of endocytic/exocytic-recycling events has proven difficult until recently because of lack of sensitivity, limited speed, photobleaching and phototoxicity associated with conventional imaging modalities. The Lattice Light Sheet Microscope (LLSM) allows to overcome these difficulties, yet reaching high spatial resolution. 3D images are captured for several minutes at a high acquisition frequency, and enables the study of signaling, transport, and stochastic self-assembly in complex environments. In addition, this imaging technique and 3D-tracking allow us to characterize the molecular machineries involved in the exocytosis and endocytosis mechanisms. We have the opportunity to observe a series of sequential events corresponding to the fusion with the plasma membrane (exocytosis) and the formation of endocytic carriers, including the trafficking of vesicles throughout the entire membrane system. We have got preliminary results of the coordination of vesicle recycling from the endosomal recycling compartment up to the plasma membrane using LLSM imaging and 3D tracking. In addition, we introduced a quantitative analysis of endocytosis dynamics of AP2 adaptor complex, Galectin-3 (see Figure 6) and Transferrin using single particle tracking analysis of 3D+time data. These case studies clearly demonstrated the advantage of lattice light sheet microscopy for imaging endocytic/exocytic events in single cells.

Software: THOTH and CPAnalysis (see Sections 6.2 and 6.6).

Collaborators: C. Wunder and L. Johannes (Institut Curie, PSL Research University, Cellular and Chemical Biology, U1143 INSERM / UMR 3666 CNRS).

7.4. 3D flow estimation in 3D fluorescence microscopy image sequences

Participants: Sandeep Manandhar, Patrick Bouthemy, Charles Kervrann.

We have proposed a variational approach for 3D optical flow computation from a pair of fluorescence microscopy volumes. This computational method has been extensively evaluated on appropriate simulated data. To simulate a volume pair with a ground truth flow field, we extended the Horn-and-Schunck optical flow method to 3D (3DHS). The computed flow field by 3DHS between two input images is then used to generate a new pair volume. The new source and target volumes along with the corresponding 3DHS flow fields serve as the dataset with the ground truth. The latter was used for the parameterization of our method and for the comparative study with the state-of-the art method proposed by Amat et al., 2012 [44].

Meanwhile, we proposed a novel error measure named SAE (for Structural Angular Error) for 3D optical flow, in absence of any ground truth flow field (see Figure 7). SAE measures the angular difference between the principal orientations of the structures present in the backward warped target and the source volumes at each voxel. We found out that the average of SAE (ASAE) and average of the end-point-error (a standard optical flow error in presence of ground truth) behave similarly.

We also integrated L_1 regularization in our variational approach. In contrast to our previous L_2 regularization approach, this method preserves discontinuities of the flow field. However, both of our methods are time demanding and not parallelizable in implementation. Then, we integrated our Census Signature-based data term with total variation regularization that also produces discontinuous flow fields. Consequently, we took the splitting approach for optimization. A gain of four times was obtained in the calculation speed with 12-core implementation of the new method, compared to our previous two methods, for still similar ASAE score.

We have also proposed two new methods for the visualization of the 3D flow fields, named 3DHSV and MIP-flow respectively. The 3DHSV method color codes the 2D projections of the flow field in the Hue and the Saturation color spaces, while mapping the off-the-plane motion to the Value space. MIP-flow also encodes the 2D projections to the Hue and the Saturation spaces. However, it only considers encoding the 2D projections of the 3D vectors corresponding to the maximum intensity points in the direction perpendicular to the projection plane. This work is carried out in collaboration with UTSW Dallas in the frame of the Inria associated team CytoDI.

Software: FlowScope (see Section 6.7).
Collaborators: P. Roudot, E. Welf and G. Danuser (UTSW Medical Center, Dallas, USA).

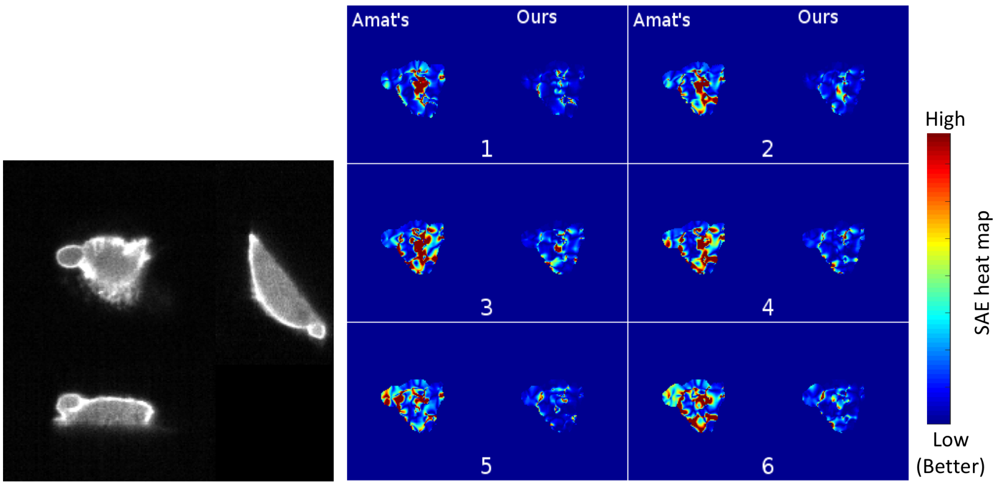


Figure 7. SAE maps (right) to compare our variational method to the Amat's method [44] applied to a 3D image pair (left) depicting a MV3 melanoma cell undergoing blebbing on a coverslip observed with Diagonally scanned Light-Sheet Microscopy (2.86 Hz sampling frequency, $300 \times 300 \times 83/50 \times 50 \times 30\mu m^3$*) (input images by courtesy of Danuser lab, UTSW Dallas, USA).*

7.5. Probabilistic overall reconstruction of membrane-associated molecular dynamics from partial observations in rod-shaped bacteria

Participants: Yunjiao Lu, Charles Kervrann.

Understanding the mechanisms that maintain the structure of rod-shaped bacteria is a challenging problem in cell biological research. Thanks to progress in molecular biology and microscopy (e.g Total Internal Reflection Fluorescence (TIRF) microscopy), we have the opportunity to observe the dynamics of the cell wall construction workers, that is the membrane-associated molecular machines (MMs). Due to the cylindrical form of the bacteria and the 2D selective visualization in TIRF microscopy, only around one third of the perimeter can be observed at a given time. Nevertheless, from the partial observed bacteria surface images, earlier studies showed that a fraction of the MMs performs directed motion, across the image field quasi-orthogonally to the cylinder axis.

Accordingly, we addressed the problem of the connection of motion segments on a cylindrical surface, assuming that one MM may re-enters into the observed region (OR), a certain period of time after having left the field of view. The directed MM motions are assumed as Brownian motion with drift. The birth and death events of the MMs are supposed to happen independently and uniformly on the surface. Given a set of observed segments entering and exiting the OR, we proposed a probabilistic framework to calculate the probabilities of the events of birth, death and re-entry, based on speed and diffusion of the motion and the time of exit and entry. Even though two third of the surface is hidden as shown in Figure 8, this framework allows us to derive a computational procedure aiming at connecting segments belonging to the same trajectory, and then recovering directed MMs dynamics on the whole surface. The performance of the method has been demonstrated on appropriate simulation data that mimics MMs dynamics observed in TIRF microscopy.

Collaborators: A. Trubuil and P. Hodara (INRA MaIAGE unit, Jouy-en-Josas),
R. Carballido-López and C. Billaudeau (INRA, UR MICALIS, Jouy-en-Josas).

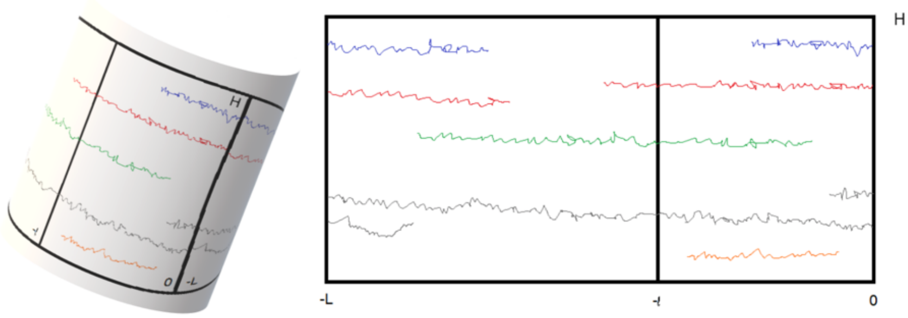


Figure 8. Trajectories on the cylinder and its 2D representation. The unobserved region is $(-L,-l] \times [0,H]$ and the observed region is $(l,0] \times [0,H]$.

7.6. Data assimilation and modeling of cell division mechanism

Participants: Anca-Georgiana Caranfil, Charles Kervrann.

Asymmetric cell division is a complex process that is not yet fully understood. A very well-known example of such a division is the first division of C.elegans embryo. To improve our understanding of this process, we used mathematical modeling to study the first division of C.elegans embryo, both on wild type cells and under a wide range of genetic perturbations. Asymmetry is clearly visible at the end of the anaphase, as the mitotic spindle is off-center. The study of the mitotic spindle dynamics is, thus, a useful tool to gain insights into the general mechanics of the system used by the cell to correctly achieve asymmetric division. The overall spindle behavior is led by the spindle poles behavior. We proposed a new dynamic model for the posterior spindle pole that explains the oscillatory behavior during anaphase and confirms some previous findings, such as the existence of a threshold number of active force-generator motors required for the onset of oscillations. We also confirmed that the monotonic increase of motor activity accounts for their build-up and die-down. By theoretically analyzing our model, we determined boundaries for the motor activity-related parameters for these oscillations to happen. This also allowed us to describe the influence of the number of motors, as well as physical parameters related to viscosity or string-like forces, on features such as the amplitude and number of oscillations. Lastly, by using a Bayesian approach to confront our model to experimental data, we were able to estimate distributions for our biological and bio-physical parameters. These results give us insights on variations in spindle behavior during anaphase in asymmetric division, and provide means of prediction for phenotypes related to misguided asymmetric division. This model will be instrumental in probing the function of yet undocumented genes involved in controlling cell division dynamics.

Collaborators: Y. Le Cunff and J. Pécréaux (IGDR – Institute of Genetics & Development of Rennes).

7.7. Convolutional Neural Networks algorithms for calcium signal segmentation in astrocytes in 3D lattice light sheet microscopy

Participants: Anais Badoual, Charles Kervrann.

Astrocytes, glial cells of the central nervous system, are detectors and regulators of neuronal information processing. It is established that neuronal synapses are physical sites of intercellular contact that transmit and transform information in a very rapid and flexible way, playing a pivotal role for learning and memory formation as well as neurological diseases of the mammalian brain. Astrocytes are thought to integrate neuronal inputs and modulate information transfer between neurons. In particular, cytoplasmic calcium signaling in astrocytes is believed to be crucial for astrocyte-neuron communication. However, quantification of intracellular calcium signals in astrocytes is hindered by the complexity of their cell shape, that consists of a cell body sprouting a highly ramified set of large to very fine protrusions called processes. Until recently, the quantification of intracellular propagation of calcium signal in astrocytes with fluorescent calcium indicators has been restricted to two dimensions, either 2D cell cultures or 2D slicing of a 3D setup. However it is not clear what amount of information is lost by ignoring the 3rd dimension in these experiments. The emergent 3D Lattice Light Sheet Microscopy (LLSM) is a powerful and promising technology (voxel size: 250nm x 250nm x 700nm; acquisition time: 200 frames per second) to give a much more complete and refined view of the dynamic behavior of calcium signaling in astrocytes inside living brain slices and in the intact mouse brain in vivo. Unfortunately, we lack image analysis tools to locate, segment, track and quantify the propagation of those 3D calcium signals in very ramified cell shapes.

In this context, we have started to develop an image processing tool for neurobiologists that 1) detects and segments calcium signals in 3D+time LLSM images, and 2) classifies these signals based on their 3D space-time morphological characterization. To do so, we focus on 3D convolutional network and machine learning techniques.

Collaborators: V. Nägerl and M. Arizono (Interdisciplinary Institute for Neuroscience, Bordeaux),
H. Berry and A. Denisot (EPC BEAGLE, Inria Rhone-Alpes).

7.8. Geo-colocalization and coorientation in fluorescence super-resolution microscopy

Participants: Frédéric Lavancier, Reda Alami Chantoufi, Aymeric Lechevranton, Antoine Salomon, Charles Kervrann.

Colocalization aims at characterizing spatial associations between two fluorescently-tagged biomolecules by quantifying the co-occurrence and correlation between the two channels acquired in fluorescence microscopy. This problem remains an open issue in diffraction-limited microscopy and raises new challenges with the emergence of super-resolution imaging. In [19], we proposed an original method (GcoPS) that exploits the random sets structure of the tagged molecules to provide an explicit testing procedure. GcoPS requires the adjustment of a p-value that guarantees more reproducibility and more objective interpretation and takes as inputs two 2D or 3D binary segmented images. This year, we extended this approach to the estimation of local co-localization. This amounts to applying the statistical test on windows randomly drawn in the whole image. A multiple testing procedure allows us to compute a global partial colocalization score. Meanwhile, the excursion sets of colocalization score map estimated by Gaussian smoothing are very helpful to detect regions of interest corresponding to significant colocalization and anti-colocalization sites. This approach has been evaluated on STORM (Stochastic Optical Reconstruction Microscopy) images which provides several hundreds thousands of super-localized positions of individual molecules with an average accuracy of 10-20 nanometers (see Figure 9). Finally, the method has been successively extended to the geo-coorientation (or geo-coalignment) of 2D-3D vectors (optical flow, tensors) and trajectories to analyze the molecular interactions.

Software: GcoPS (see Section 6.1).

Collaborators: J. Salamero (CNRS-UMR 144, Institut Curie, PSL Research University),
G. Bertolin (IGDR – Institute of Genetics & Development of Rennes),
M. Lelek and C. Zimmer (Institut Pasteur, Paris).

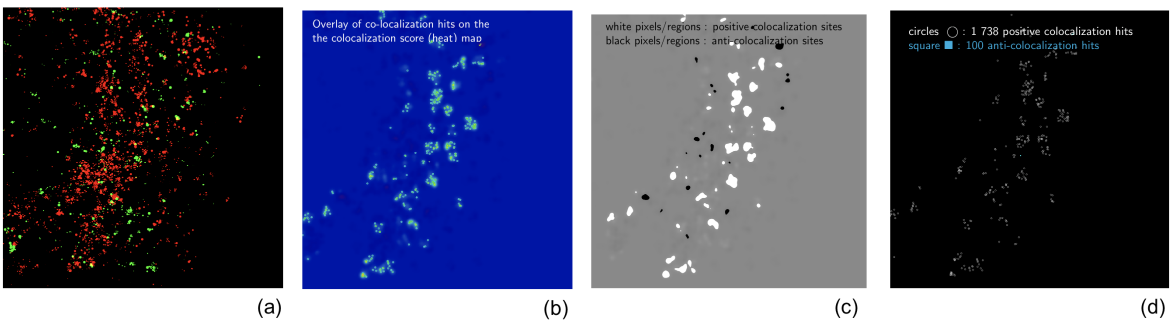


Figure 9. Illustration of geo-colocalization of two molecules (red/green channels) in STORM super-resolution microscopy (original image size: 4576 $\times$ 3564 pixels; pixel size: 3 nanometers). (a) Overlay of two channels (sub-region of the original pair); (b) colocalization hits overlaid on the score (heat) map; (c) excursion sets of detected colocalization (white) and anti-colocalization (black) sites overlaid on the score map; (d) detected co-localization (circles) and anti-co-localization (squares) hits.

7.9. Immersive and interactive visualization of 3D temporal data using a space time hypercube

Participants: Gwendal Fouché, Charles Kervrann.

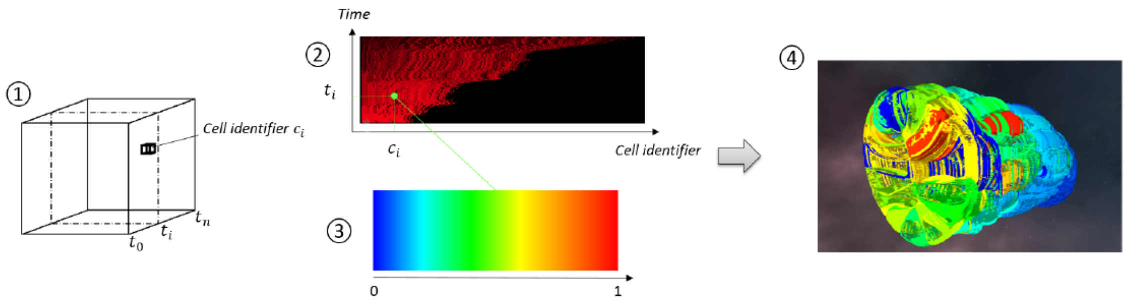


Figure 10. Flow diagram of the STC generation. In step 1, the user places the interactive clipping plan to get the desired cross-section. In step 2, camera parameters are automatically set in order to render the cross section at each time point. In step 3, the top image presents the output of the rendering operation, using the RGB channel to save cell identifiers; the bottom image is the result of an edge detection filter that will be useful for display. In step 4, the rendered images are stacked into a 3D texture. Each voxel contains a cell identifier, and its position in terms of depth indicates a time point t_i.

The analysis of multidimensional time-varying datasets, whose size grows as recording and simulating techniques advance, faces challenges on the representation and visualization of dense data, as well as on the study of temporal variations. In this context, we proposed an extension of the well-known Space-Time Cube (STC) visualization technique in order to visualize time-varying 3D spatial data acquired in 3D fluorescence microscopy, taking advantage of the interaction capabilities of Virtual Reality. The extended STC is based on a user-driven projection of the spatial and temporal information modeled as a 4D Space-Time Hypercube (STH). This projection yields a 3D STC visualization, which can also encode non-spatial quantitative data. Moreover, we proposed a set of tools allowing the user to manipulate the 3D STC that benefits from the visualization, exploration and interaction possibilities offered by immersive environments (see Figure 10). Finally, the extended STC has been integrated in a VR application for visualization of spatiotemporal biological data, illustrating the usage of the proposed visualization method for the morphogenesis analysis.

Collaborators: F. Argelaguet (EPC HYBRID, Inria Rennes),
E. Faure (Laboratory of Computer Science, Robotics and Microelectronics of Montpellier).

7.10. Unsupervised motion saliency map estimation based on optical flow inpainting

Participants: Léo Maczyta, Patrick Bouthemy.

We have addressed the problem of motion saliency in videos. Salient moving regions are regions that exhibit motion departing from their spatial context in the image, that is, different from the surrounding motion. In contrast to video saliency approaches, we estimate dynamic saliency based on motion information only. We propose a new unsupervised paradigm to compute motion saliency maps. The key ingredient is the flow inpainting stage. We have to compare the flow field in a given area, likely to be a salient moving element, with the flow field that would have been induced in the very same area with the surrounding motion. The former can be computed by any optical flow method. The latter is not directly available, since it is not observed. Yet, it can be predicted by a flow inpainting method. This is precisely the originality of our motion saliency approach.

Our method is then two-fold. First, we extract candidate salient regions from the optical flow boundaries. Secondly, we estimate the inpainted flow using an extension of a diffusion-based method for image inpainting, and we compare the inpainted flow to the original optical flow in these regions. We interpret the possible discrepancy (or residual flow) between the two flows as an indicator of motion saliency. In addition, we combine a backward and forward processing of the video sequence. The method is flexible and general enough, by relying on motion information only. Experimental results on the DAVIS 2016 benchmark demonstrate that the method compares favorably with state-of-the-art video saliency methods. Additionally, by estimating the residual flow, we provide additional information regarding motion saliency that could be further exploited (see Figure 11).

Collaborators: O. Le Meur (Percept team, IRISA, Rennes).

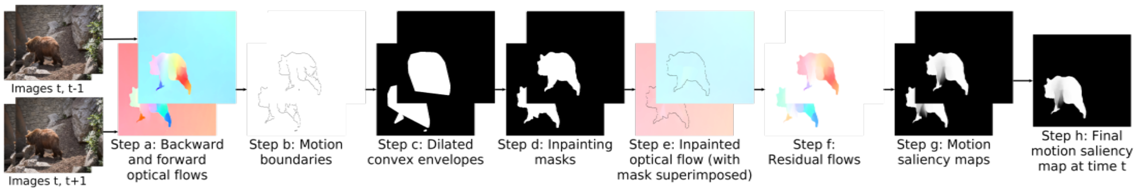


Figure 11. Workflow of the motion saliency estimation method.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral grants with industry

8.1.1. Contract with Fourmentin-Guilbert Foundation: Macromolecule detection in 3D cellular cryo-electron tomograms

Participants: Emmanuel Moebel, Charles Kervrann.
Duration: 5 months (Dec 2019 – Apr 2020).

The objective of the project is to improve the DeepFinder software dedicated to the detection and identification of macromolecules within 3D cellular cryo-electron tomograms. In collaboration with Fourmentin-Guilbert Foundation, the goal is to build cellular atlases of several organisms from localizations of macromolecules (see Software DeepFinder in Section 6.9).

Funding: Fourmentin-Guilbert Foundation.
Collaborators: D. Larivière & E. Fourmentin (Fourmentin-Guilbert Foundation), A. Martinez & W. Baumeister (Max Planck Institute, Martinsried, Germany).

8.1.2. Contract with DGA: Motion saliency analysis in videos

Participants: Léo Maczyta, Patrick Bouthemy.
Duration: 36 months (Oct 2017 – Sep 2020).

This project funded by the DGA (Ministry of defense) and Région-Betagne concerns the PhD thesis (co-funding) carried out by Léo Maczyta. The goal is to develop motion saliency methods along three axes: temporal motion saliency detection, saliency map estimation, trajectory-based saliency detection (see Section 7.10).

Funding: DGA (National Defense Agency) and Région-Bretagne.

8.1.3. Contract with GATACA Systems: Super-resolution microscopy and in live cell imaging

Participants: Jean Salamero, Ludovic Leconte, Charles Kervrann.
Duration: 36 months (Jan 2017 – Dec 2019).

The objective of the project is to transfer innovations for Multi-Angle TIRFM (using Azymuthal TIRFM from Ilas2) and collaborate as "β-Test site" for SIM in Nipkow disk microscopy (product: Live-SR).

Funding: GATACA Systems company.
Collaborators: C. Gueudry (GATACA Systems), J. Boulanger (MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, UK).

8.1.4. Contract with CryoCapCell SA: 3D LIVE CLEM (Correlative Light and Electron Microscopy) to decipher fates and functions of exosomes in vivo

Participant: Jean Salamero.
Duration: 24 months (Oct 2018 – Sep 2020).

The objective of the project is to link dynamic biogenesis of intracellular membrane compartments with their ultrastructures. It combines fast high resolution photonic imaging (MA-TIRFM and fast high pressure freezing for 3D cryoEM. It requires adapted registration methods in 3D, in order to navigate through the multiple scales.

Funding: DIM–ELICIT Empowering LIfe sCiences with Innovative Technologies (Région Ile de France).
Collaborators: G. Van Niel (coordinator, Institute of Psychiatry and Neuroscience of Paris), G. Raposo (CNRS-UMR 144 Institut Curie PSL Research), X. Heiligenstein (CryoCapCell SA).

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Motion saliency analysis in videos

Participants: Léo Maczyta, Patrick Bouthemy.
Duration: 36 months (Oct 2017 – Sep 2020).

See Section 8.1.2.

Funding: DGA (National Defense Agency) and Région-Bretagne.

9.2. National Initiatives

9.2.1. France-BioImaging project

Participants: Sylvain Prigent, Patrick Bouthemy, Charles Kervrann, Jean Salamero.
Duration: 2011 – 2024.

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 co-head 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 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) (see Section 6.13).

Funding: Investissement d'Avenir, ANR INBS-PIA 2011.
Coordinator: CNRS (J. Salamero, UMS 3714 CEMIBIO & CNRS-UMR 144, Institut Curie, PSL Research University).
Partners: CNRS, University of Paris-Diderot-Paris 7, Aix-Marseille University, University of Bordeaux, University of Montpellier, Institut Pasteur, Institut Curie, Inria, ENS Paris, University of Paris Descartes, UPMC, Ecole Polytechnique, INSERM.

9.2.2. ANR NucleoPLASTIC: Plasticity of the Nuclear Pore Complex

Participant: Jean Salamero.
Duration: 48 months (Oct 2015 – Sep 2019).

In this project, we have deciphered molecular/structural changes on the nuclear face of the Nuclear Pore Complex, their dynamics during cell division, and highlighted their role in the dynamics of association with the heart of the pore with consequences on maintaining the integrity of the genome. This was possible through the development of a 3D localization software GenLoc3D (https://team.inria.fr/serpico/software/genloc3d/, FIJI/ImageJ plug-in).

Funding: ANR (Agence Nationale de la Recherche).
Coordinator: C. Dargemont (INSERM, Hopital St Louis, Paris).
Partners: CNRS-UMR 144, Institut Curie, PSL Research, Paris.

9.2.3. ANR DALLISH project: Data Assimilation and Lattice LIght SHeet imaging for endocytosis/exocytosis pathway modeling in the whole cell

Participants: Antoine Salomon, Anca-Georgiana Caranfil, Sandeep Manandhar, Cesar Augusto Valades Cruz, Patrick Bouthemy, Ludovic Leconte, Jean Salamero, Charles Kervrann.
Duration: 48 months (Oct 2016 – Sep 2020).

Cutting-edge Light Lattice Sheet 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 motion estimation and tracking. 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 (see Sections 7.2–7.6 and 7.8).

Funding: ANR (Agence Nationale de la Recherche) PRC (Collaborative Research Project).
Coordinator: C. Kervrann.

Partners: Inria (SERPICO, BEAGLE, FLUMINANCE teams), INRA MaIAGE Unit Jouy-en-Josas, Institut Curie (CNRS-UMR 144 & U1143 INSERM / UMR 3666) Paris.

9.2.4. Inria Project Labs (IPL / DEFI), Exploratory Research Actions and Technological Development Actions

9.2.4.1. NAVISCOPE: image-guided NAvigation and VISualization of large data sets in live cell imaging and microCOPy

Participants: Gwendal Fouché, Cesar Augusto Valades Cruz, Ludovic Leconte, Anais Badoual, Jean Salamero, Charles Kervrann.
Duration: 60 months (2018 – 2022).

In the frame of the "Naviscope" IPL project (https://project.inria.fr/naviscope/), our objective is to develop original and cutting-edge visualization and navigation methods to assist scientists, enabling semi-automatic analysis, manipulation, and investigation of temporal series of multi-valued volumetric images, with a strong focus on live cell imaging and microscopy application domains. Naviscope, built upon the strength of scientific visualization and machine learning methods, will provide systems capable to assist the scientist to obtain a better understanding of massive amounts of information. Such systems will be able to recognize and highlight the most informative regions of the dataset by reducing the amount of information displayed and guiding the observer attention. We address the three following challenges and issues:

- Novel machine learning methods able to detect the main regions of interest, and automatic quantification of sparse sets of molecular interactions and cell processes during navigation to save memory and computational resources.
- Novel visualization methods able to encode 3D motion and deformation vectors and dynamics features with color and texture-based and non-sub-resolved representations, abstractions, and discretization, as used to display 2D motion and deformation vectors and patterns.
- Effective machine learning-driven navigation and interaction techniques for complex functional 3D+Time data enabling the analysis of sparse sets of localized intra-cellular events and cell processes (migration, division, etc.) (see Section 7.9).

Meanwhile, we address the technological challenge of gathering up the software developed in each team to provide a unique original tool for users in biological imaging, and potentially in medical imaging.

Funding: Inria (IPL / DEFI).
Coordinator: C. Kervrann.
Partners: AVIZ Inria team (Saclay); BEAGLE Inria team (Lyon), HYBRID Inria team (Rennes), MORPHEME Inria team (Sophia-Antipolis); MOSAIC Inria team (Lyon), PARIETAL Inria team (Saclay), SERPICO Inria team (Rennes); MaIAGE INRA Unit (Jouy-en-Josas); CNRS-UMR 144, Institut Curie, PSL Research University (Paris).

9.3. European Initiatives

9.3.1. Collaborations in European Programs, Except FP7 & H2020

9.3.1.1. ESFRI initiative programm

SERPICO is involved in the ESFRI Euro-BioImaging (https://www.eurobioimaging.eu/) initiative, 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 is also involved in the French initiative, 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: Turku University (J. Eriksson, Turku, Finland).
Funding: Member states of the European Union.
Partners: 15 European countries.

9.3.1.2. EIT Digital program

Participants: Sylvain Prigent, Charles Kervrann.
Duration: 12 months (Nov 2019 – Oct 2020).

SERPICO is involved in a European project which aims at developing a connected wearable device for diagnosis and treatment of photodermatoses. Using the data on skin sun sensitivity and UV exposure habits with machine learning algorithms will enable to make more precise optimal sun exposure predictions for patients. The wearable device will be useful for a larger population to increase awareness around overexposure to UV as a main cause of sun damage and worst-case skin cancer.

Funding: EIT Digital.
Inria coordinator: C. Kervrann.
Partners: UVisio and Nobleo Projects B.V., Eindhoven, The Netherlands.

9.4. International Initiatives

9.4.1. Inria International Partners

9.4.1.1. Informal International Partners

- Collaboration with Max-Planck Institute, Martinsried, Germany (with A. Martinez and W. Baumeister): Detection and segmentation of macromolecules in cryo-electron tomography (project in progress with E. Moebel and C. Kervrann) (see Sections 6.9 [30] and 6.12 [22]).
- Collaboration with University of Texas SouthWestern (UTSW) Medical Center, Dallas, United States (P. Roudot, E. Welf and G. Gaudenz): 3D optical flow for cell migration quantification (project in progress with S. Manandhar, P. Bouthemy and C. Kervrann) (see Sections 6.11 and 7.4 [21]).
- Collaborations with the MRC laboratory of Molecular Biology (with E. Derivery and J. Boulanger) and the Cambridge Advanced Imaging Centre (with L. Muresan), Cambridge, UK (project in progress with A. Salomon and C. Kervrann) (see Section 7.2).
- Collaboration with the PKU University, Institute of Molecular Medicine, Beijing (with L. Chen and Y.M. Liu): 3D reconstitution of the biogenesis of Endoplasmic Reticulum-plasma membrane contact sites (ER-PM MSCs upon Ca2+ store depletion or replenishment) (project in progress with C.A. Valades Cruz and J. Salamero).

9.5. International Research Visitors

9.5.1. Visits to International Teams

- Charles Kervrann visited the MRC laboratory of Molecular Biology and the Cambridge Advanced Imaging Centre (June, 1 week, Cambridge, UK).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organization

10.1.1.1. General Chair, Scientific Chair

- Charles Kervrann was co-General Chair (with R. Ober, Texas A&M), and head of the organizing committee of the “Quantitative BioImaging” (QBI’2019) conference (350 attendees, Rennes, January 2019).
- Jean Salamero was the scientific responsible and co-organizer of the first joint "Workshop on BioImaging” between Paris Sciences Lettres (PSL-Institut Curie) University and Peking University (PKU) (MoU for official “collaboration and exchange program PSL Qlife/PKU in the Bioimaging field 2020-2024).

10.1.2. Scientific Events: Selection

10.1.2.1. Member of the Conference Program Committees

- Charles Kervrann: member of the scientific committee of the international "Quantitative BioImaging” (QBI’2019, QBI’2020) conference, member of the scientific committee of the international NEUBIAS 2020 conference, and member of the scientific committee of the JIONC workshop (GdR Ondes, since 2014).
- Patrick Bouthemy: Associate Editor for the ISBI’2019 conference.

10.1.2.2. Reviewer

- Charles Kervrann: reviewer for ISBI’2019.

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 Computer Science", section "Computer Vision and Image Analysis".

10.1.3.2. Reviewer - Reviewing Activities

- Charles Kervrann: Journal of Mathematical Imaging and Vision, IEEE Signal Processing Letters, Nature Communications, IEEE Transactions on Medical Imaging, OSA Biomedical Optics Express.
- Patrick Bouthemy: IEEE Signal Processing Letters.
- Jean Salamero: Nature Methods, Journal of Cell Sciences, Nature Communications, Biology of the Cell, ACS Photonics, Journal of Microscopy.

10.1.4. Invited Talks

Charles Kervrann:

- "Statistical methods for intracellular dynamics classification in live cell imaging", "Bioimage Processing" workshop, June, Cambridge, UK (see [29]).
- "Computational methods for intracellular dynamics in live cell imaging", “Random Walks and Intracellular Transport” workshop, Apr, Manchester, UK (see [28]).
- "A fast statistical colocalization method for 3D live cell imaging and super-resolution microscopy", "Statistical Modeling for Shapes and Imaging” workshop, March, Paris (see [27]).
- "Geolocalization and classification of macromolecules in cryo-electron microscopy with deep-learning", Journées Scientifiques de l’Université de Nantes: "la microscopie électronique pour les sciences du vivant", June, Nantes.
- "Computational methods for fluorescence microscopy", Institut Curie CNRS-UMR 168 Seminar, February, Paris.

Patrick Bouthemy:

- "Intensity-based methods for fully automated registration in 2D and 3D CLEM", COMULIS & BioImaging Austria/CMI conference, November, Vienna, Austria (see [26]).

Jean Salamero:

- "Live cell imaging in High Space-time Resolution" and BioImaging course (Co-Organizer), ReTUBI H2020) Super Resolution meeting and Imaging, October, Lisbon, Portugal.
- "Career Path for the Next Generation of BioImaging Scientist: Training Passport", Global BioImaging 2019 "EOE.V", September, Singapore.

Sylvain Prigent:

- "Portail analyse d'images France-BioImaging", 16èmes Assises Nationales des Plateformes du réseau RT-mfm, March, Rennes.
- "Portail analyse d'images France-BioImaging", 9èmes Journées Scientifiques et Techniques du Réseau des Microscopistes de l'INRA, November, Nantes.

10.1.5. Leadership within the Scientific Community

- Charles Kervrann is member of the executive board of the GdR MIV/ImaBio (2588 - Microscopie Fonctionnelle du Vivant) CNRS and member of the scientific committee of the Interdisciplinary MiFoBio School CNRS (http://www.mifobio.fr).
- 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 member of the selection committee for an Assistant-Professor (Maitre de Conférences) position at the University of Paris-Descartes (Section CNU 26, Mathematics). He was reviewer for European Research Council (ERC Consolidator Grant).
- Patrick Bouthemy was member of the committee for professor promotion at IMT Atlantique.
- Jean Salamero was reviewer for European Research Council (Junior Grant), ANR and ITMO Cancer program.

10.1.7. Research Administration

Charles Kervrann:

- Member of the executive board of the project committee ("bureau du Comite´ des Projets") of the Inria Rennes - Bretagne Atlantique centre since 2010.
- Co-head of the "BioImage Informatics" node (ANR France-BioImaging project http://france-bioimaging.org/), National Research Infrastructure for Biology and Health) since 2011.

Patrick Bouthemy:

- Head of Excellence Lab (Labex) CominLabs (http://www.cominlabs.ueb.eu) since April 2014.
- 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/).
- Inria representative in the steering committee of the DGA-Inria collaboration.
- Member of the Research Committee of IMT Atlantique.
- Head of the evaluation committee of the EORD profession of the DGA and head of the HCERES evaluation committee of the DTIS department of ONERA.
- Member of the ANR evaluation committee for the call for proposals regarding the Artificial Intelligence field.

Jean Salamero:

- Director of the National Research Infrastructure in Biology and Health, France BioImaging (2015-2020) (https://france-bioimaging.org/).
- Director of the Service Unit UMS 3714 (2015-2020) CEMIBIO CNRS-Institut Curie.
- Member of the Advisory Committee for the creation of the National Biomedical Imaging Center (NBIC@Beijing, Republic of China) since 2017.
- Member of the Advisory Board of EMBL Core Facilities.
- Member of the Steering Committee of the DIM-ELICIT program (Région Ile de France) since 2016.
- Member of the Qlife "Institut Convergences" since 2018.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Charles Kervrann:

- Engineer Degree: Genomics and Informatics, 4.5 hours, MINES ParisTech.
- 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: Analysis of Image Sequences, 9 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.

10.2.2. Supervision

- Anca-Georgiana Caranfil (PhD in progress): Data assimilation methods for cell division mechanisms and molecule trafficking analysis (started in December 2016, supervised by C. Kervrann and Y. Le Cunff (IGDR, Rennes)).
- Gwendal Fouché (PhD in progress): Immersive interaction and visualization of temporal 3D data (started in October 2019, supervised by C. Kervrann, F. Argelaguet (EPC HYBRID, Inria, Rennes), and E. Faure (LIRMM, Montpellier)).
- Yunjiao Lu (PhD in progress): Intracellular dynamics and super-resolution imaging: analysis of bacteria wall at the molecular scale (started in October 2017, supervised by C. Kervrann, A. Trubuil and R. Carballido-Lopez (INRA, Jouy-en-Josas)).
- Léo Maczyta (PhD in progress): Motion saliency in video sequences (started in October 2017, supervised by P. Bouthemy and O. Lemeur (Team Percept, IRISA, Rennes)).
- Sandeep Manandhar (PhD defended [11]): Optical flow methods for 3D fluorescence imaging (defended in November 2019, supervised by P. Bouthemy and C. Kervrann).
- Emmanuel Moebel (PhD defended [12]): New strategies for the identification and enumeration of macromolecules in 3D images of cryo electron tomography (defended in January 2019, supervised C. Kervrann).
- Antoine Salomon (PhD in progress): Statistical aggregation for image analysis in fluorescence microscopy and super-resolution (started in November 2017, supervised by C. Kervrann).

10.2.3. Juries

Charles Kervrann:

- Reviewer of the HdR of E. Baudrier (University of Strasbourg, ICube – UMR 7357).

- Reviewer of the PhDs of A. Davy (University of Paris Saclay, Ecole Normale Supérieure Paris-Saclay, supervised by J.-M. Morel and A. Desolneux), P. Parutto (University of Paris Sciences et Lettres, Ecole Normale Supérieure, supervised by D. Holcman), J.D.K. Hansen (University of Copenhagen, DIKU, Denmark, supervised by F. Lauze).
- President of PhD thesis committee of M. Tassano (University of Paris-Descartes, supervised by J. Delon).
- Member of the PhD committee of P. Naylor (University of Paris Sciences et Lettres, MINES ParisTech, supervised by T. Walter).
- External examiner of the application of M. Arigovindan (Faculty assessment, promotion to the position of Associate Professor with tenure, Indian Institute of Science).

Patrick Bouthemy:

- President of the PhD thesis committee of A. Lopez (University of Rennes 1, supervised by J. Pettré and F. Chaumette).

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^o 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^o 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^o 10 [*DOI :* 10.1371/journal.pone.0013190]

[4] 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^o 2, p. 442-453 [*DOI :* 10.1109/TMI.2009.2033991]

[5] V. Briane, C. Kervrann, M. Vimond.*Statistical analysis of particle trajectories in living cells*, in "Physical Review E ", June 2018, vol. 97, n^o 6, p. 1-20 [*DOI :* 10.1103/PhysRevE.97.062121], https://hal.inria.fr/hal-01961971

[6] 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 four-dimensional optical microscopy*, in "Proc Natl Acad Sci USA", Sep 2010, vol. 107, n^o 37, p. 16016-16022 [*DOI :* 10.1073/pnas.1004037107]

[7] 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^o 4, p. 2484-2520 [*DOI :* 10.1137/120872048], https://hal.inria.fr/hal-00931667

[8] 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

[9] 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^o 2, 14 [*DOI :* 10.1109/TIP.2014.2380178], https://hal.inria.fr/hal-01103126

[10] T. PÉCOT, C. KERVRANN, J. SALAMERO, J. BOULANGER.*Counting-based particle flux estimation for traffic analysis in live cell imaging*, in "IEEE Journal of Selected Topics in Signal Processing", September 2015 [*DOI :* 10.1109/JSTSP.2015.2482460], https://hal.inria.fr/hal-01244946

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] S. MANANDHAR.*3D Motion Estimation and Assessment in Fluorescence Microscopy Volume Sequences*, Université de Rennes 1 [UR1], November 2019, https://hal.archives-ouvertes.fr/tel-02418551

[12] E. MOEBEL.*New strategies for the identification and enumeration of macromolecules in 3D images of cryo electron tomography*, Université de Rennes 1, February 2019, https://hal.inria.fr/tel-02153877

Articles in International Peer-Reviewed Journal

[13] P. BOUTHEMY, B. M. TOLEDO ACOSTA, B. DELYON.*Robust model selection in 2D parametric motion estimation*, in "Journal of Mathematical Imaging and Vision", September 2019, vol. 61, n[o] 7, p. 1022-1036 [*DOI :* 10.1007/S10851-019-00883-2], https://hal.archives-ouvertes.fr/hal-02315977

[14] V. BRIANE, A. SALOMON, M. VIMOND, C. KERVRANN.*A computational approach for detecting microdomains and confinement domains in cells: a simulation study*, in "Physical Biology", December 2019, p. 1-26, forthcoming [*DOI :* 10.1088/1478-3975/AB5E1D], https://hal.inria.fr/hal-02424860

[15] V. BRIANE, M. VIMOND, C. KERVRANN.*An overview of diffusion models for intracellular dynamics analysis*, in "Briefings in Bioinformatics", June 2019, p. 1-15 [*DOI :* 10.1093/BIB/BBZ052], https://hal.inria.fr/hal-02424761

[16] V. BRIANE, M. VIMOND, C. A. VALADÉS-CRUZ, A. SALOMON, C. WUNDER, C. KERVRANN.*A sequential algorithm to detect diffusion switching along intracellular particle trajectories*, in "Bioinformatics", June 2019, p. 1-13 [*DOI :* 10.1093/BIOINFORMATICS/BTZ489], https://hal.inria.fr/hal-02424777

[17] B. COMBÈS, S. PRIMA.*An efficient EM-ICP algorithm for non-linear registration of large 3D point sets*, in "Computer Vision and Image Understanding", November 2019, p. 1-14 [*DOI :* 10.1016/J.CVIU.2019.102854], https://hal.inria.fr/hal-02414127

[18] F. LAVANCIER, A. POINAS, R. WAAGEPETERSEN.*Adaptive estimating function inference for non-stationary determinantal point processes*, in "Scandinavian Journal of Statistics", 2019, p. 1-35, forthcoming, https://hal.archives-ouvertes.fr/hal-01816528

[19] F. LAVANCIER, T. PÉCOT, L. ZENGZHEN, C. KERVRANN.*Testing independence between two random sets for the analysis of colocalization in bioimaging*, in "Biometrics", August 2019 [*DOI :* 10.1111/BIOM.13115], https://hal.archives-ouvertes.fr/hal-02369555

[20] L. MACZYTA, P. BOUTHEMY, O. LE MEUR.*CNN-based temporal detection of motion saliency in videos*, in "Pattern Recognition Letters", December 2019, vol. 128, p. 298-305 [*DOI :* 10.1016/J.PATREC.2019.09.016], https://hal.inria.fr/hal-02345209

[21] S. MANANDHAR, P. BOUTHEMY, E. WELF, G. DANUSER, P. ROUDOT, C. KERVRANN.*3D Flow Field Estimation and Assessment for Live Cell Fluorescence Microscopy*, in "Bioinformatics", October 2019, p. 1-19, forthcoming [*DOI :* 10.1093/BIOINFORMATICS/BTZ780], https://hal.archives-ouvertes.fr/hal-02308001

[22] E. MOEBEL, C. KERVRANN.*A Monte Carlo framework for missing wedge restoration and noise removal in cryo-electron tomography*, in "Journal of Structural Biology: X", October 2019, p. 1-18, Free Open Access [*DOI :* 10.1016/J.YJSBX.2019.100013], https://hal.inria.fr/hal-02424804

[23] A. POINAS, B. DELYON, F. LAVANCIER.*Mixing properties and central limit theorem for associated point processes*, in "Bernoulli", 2019, vol. 25, n^{o} 3, p. 1724-1754 [*DOI :* 10.3150/18-BEJ1033], https://hal.archives-ouvertes.fr/hal-01519096

[24] G. POTIER, R. CAPOULADE, J. MÉROT, F. LAVANCIER, X. HEILIGENSTEIN, J. SALAMERO, P. PAUL-GILLOTEAUX.*Automatic Registration of Correlative Microscopies with Error Assessment and Applications for the Optimization of Multimodal Acquisitions*, in "Microscopy and Microanalysis", August 2019, vol. 25, n^{o} S2, p. 1020-1021 [*DOI :* 10.1017/S143192761900583X], https://hal.archives-ouvertes.fr/hal-02267554

[25] S. RAMAN, S. SINGH, T. PÉCOT, E. CASERTA, K. HUANG, J. RITTSCHER, G. LEONE, R. MACHIRAJU.*Capturing variations in nuclear phenotypes*, in "Journal of computational science", September 2019, vol. 36, p. 1-12 [*DOI :* 10.1016/J.JOCS.2019.07.001], https://hal.inria.fr/hal-02424723

Invited Conferences

[26] P. BOUTHEMY.*Intensity-based methods for fully automated registration in 2D and 3D CLEM*, in "CMI 2019 - Correlated Multimodal Imaging Conference", Vienne, Austria, November 2019, https://hal.inria.fr/hal-02419522

[27] C. KERVRANN.*A fast statistical colocalization method for 3D live cell imaging and super-resolutionmicroscopy*, in ""Statistical modeling for Shapes and Imaging" workshop", Paris, France, March 2019, https://hal.inria.fr/hal-02424596

[28] C. KERVRANN.*Computational methods for intracellular dynamics in live cell imaging*, in "“Random Walks and Intracellular Transport” workshop", Manchester, United Kingdom, April 2019, https://hal.inria.fr/hal-02424594

[29] C. KERVRANN.*Statistical methods for intracellular dynamics classification in live cell imaging*, in "“Bioimage Processing” workshop", Cambridge, United Kingdom, June 2019, https://hal.inria.fr/hal-02424593

International Conferences with Proceedings

[30] I. GUBINS, G. VAN DER SCHOT, R. VELTKAMP, F. FORSTER, X. DU, X. ZENG, Z. ZHU, L. CHANG, M. XU, E. MOEBEL, A. MARTINEZ-SANCHEZ, C. KERVRANN, T. M. LAI, X. HAN, G. TERASHI, D.

KIHARA, B. A. HIMES, J. ZHANG, S. GAO, Y. HAO, Z. LV, X. WAN, Z. YANG, Z. DING, X. CUI, F. ZHANG.*Classification in cryo-electron tomograms*, in "Eurographics Workshop on 3D Object Retrieval", Genova, Italy, May 2019 [*DOI :* 10.2312/3DOR.20191061], https://hal.inria.fr/hal-02424712

[31] L. MACZYTA, P. BOUTHEMY, O. LE MEUR.*Unsupervised motion saliency map estimation based on optical flow inpainting*, in "ICIP 2019 - IEEE International Conference on Image Processing", Taipei, Taiwan, IEEE, September 2019, p. 4469-4473 [*DOI :* 10.1109/ICIP.2019.8803542], https://hal.inria.fr/hal-02065112

Conferences without Proceedings

[32] H. BOUVRAIS, D. FAIRBRASS, L. CHESNEAU, N. SOLER, Y. LE CUNFF, T. PÉCOT, C. KERVRANN, J. PÉCRÉAUX.*Two dynamic behaviours of the microtubules at the cell cortex reveal the pulling and pushing forces that position the mitotic spondle in C. elegans embryos*, in "QBI 2019 - Quantitative BioImaging Conference", Rennes, France, January 2019, https://hal.inria.fr/hal-02424688

[33] A. CARANFIL, C. KERVRANN, Y. LE CUNFF, J. PECREAUX.*Studying oscillatory behavior in asymmetric division of Caenorhabditis elegans embryo with fluorescence microscopy*, in "QBI 2019 - Quantitative BioImaging Conference", Rennes, France, January 2019, https://hal.inria.fr/hal-02424686

[34] A. CARANFIL, Y. LE LE CUNFF, C. KERVRANN, J. PÉCRÉAUX.*Modelling oscillatory behavior in asymmetric division of C. elegans embryo*, in "BioHazard – Stochastic Models for Biology – Conference", Rennes, France, August 2019, https://hal.inria.fr/hal-02424703

[35] Y. LU, C. BILLAUDEAU, R. CARBALLLIDO-LOPEZ, C. KERVRANN, A. TRUBUIL.*3D stochastic process simulation for better interpretation of molecular dynamics related to cell wall biogenesis observed with TIRF microscopy*, in "QBI 2019 - Quantitative BioImaging Conference", Rennes, France, January 2019, https://hal.inria.fr/hal-02424699

[36] Y. LU, P. HODARA, C. KERVRANN, A. TRUBUIL.*Probabilistic overall reconstruction of membrane-associated molecular dynamics from partial observations in rod-shaped bacteria*, in ""Seeing is Believing" Conference", Heidelberg, Germany, October 2019, https://hal.inria.fr/hal-02424598

[37] S. MANANDHAR, P. BOUTHEMY, P. ROUDOT, C. KERVRANN.*3D motion estimation in 3D light microscopy image sequences: application to cell migration*, in "QBI 2019 - Quantitative BioImaging Conference", Rennes, France, January 2019, https://hal.inria.fr/hal-02424694

[38] E. MOEBEL, C. KERVRANN.*3D ConvNets improve macromolecule localization in 3D cellular cryo-electron tomograms*, in "Quantitative BioImaging (QBI) Conference", Rennes, France, January 2019, vol. 2, https://hal.inria.fr/hal-02424696

[39] A. SALOMON, V. BRIANE, M. VIMOND, C. A. VALADÉS-CRUZ, C. KERVRANN.*Analysing and quantifying intracellular particles movements in 3D LLSM data*, in "QBI 2019 - Quantitative BioImaging Conference", Rennes, France, January 2019, https://hal.inria.fr/hal-02424609

[40] C. A. VALADÉS-CRUZ, A. FORRESTER, C. WUNDER, C. KERVRANN, J. SALAMERO, L. JOHANNES.*3D tracking of endocytic events using lattice light sheet microscopy*, in "QBI 2019 - Quantitative BioImaging Conference", Rennes, France, January 2019, https://hal.inria.fr/hal-02424697

Other Publications

[41] H. BOUVRAIS, L. CHESNEAU, D. FAIRBRASS, Y. LE CUNFF, N. SOLER, T. PÉCOT, C. KERVRANN, J. PECREAUX.*Two dynamical behaviours of the microtubules at cell cortex reveal pulling and pushing forces that position the spindle in C. elegans embryo*, May 2019, working paper or preprint, https://hal.inria.fr/hal-02424717

[42] X. LE GOFF, J. COMELLES, C. KERVRANN, D. RIVELINE.*Ends and middle: global force balance determines septum location in fission yeast*, December 2019, working paper or preprint [*DOI :* 10.1101/520007], https://hal.archives-ouvertes.fr/hal-02398985

[43] Y. LU, P. HODARA, C. KERVRANN, A. TRUBUIL.*An inverse problem approach to the probabilistic reconstruction of particle tracks on a censored and closed surface*, October 2019, working paper or preprint, https://hal.inria.fr/hal-02323939

References in notes

[44] F. AMAT, E. MYERS, P. KELLER.*Fast and robust optical flow for time-lapse microscopy using super-voxel*, in "Bioinformatics", 2012, vol. 29, n^{o} 3, p. 373–380

[45] D. AXELROD.*Total Internal Reflection Fluorescent Microscopy in cell biology*, in "Traffic", 2004, vol. 2, p. 4658–4668

[46] E. BETZIG, G. PATTERSON, R. SOUGRAT, O. LINDWASSER, S. OLENYCH, J. BONIFACINO, M. DAVIDSON, J. LIPPINCOTT-SCHWARTZ, H. HESS.*Imaging intracelluar fluorescent proteins at nanometer resolution*, in "Science", 2006, vol. 313, p. 1642–1645

[47] B. CHEN ET AL.*Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution*, in "Science", 2014, vol. 346, n^{o} 6208, 1257998

[48] 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

[49] 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^{o} 3, p. 295–316

[50] N. HOZE, D. NAIR, E. HOSY, C. SIEBEN, S. MANLEY, A. HERRMANN, J.-B. SIBARITA, D. CHOQUET, D. HOLCMAN.*Heterogeneity of AMPA receptor trafficking and molecular interactions revealed by superresolution analysis of live cell imaging*, in "Proc. Natl. Academy Sciences", 2012, vol. 10, n^{o} 42, p. 17052-17057

[51] 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

[52] 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^{o} 29, p. 11705–11710 [*DOI :* 10.1073/PNAS.1121583109], http://hal.inria.fr/inserm-00717415

Project-Team SIMSMART

SIMulating Stochastic Models with pARTicles

IN COLLABORATION WITH: Institut de recherche mathématique de Rennes (IRMAR)

IN PARTNERSHIP WITH:
CNRS

Université Rennes 1

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Stochastic approaches

Table of contents

Project-Team SIMSMART

Creation of the Team: 2018 January 01, updated into Project-Team: 2019 January 01

Keywords:

Computer Science and Digital Science:

A6. - Modeling, simulation and control
A6.1. - Methods in mathematical modeling
A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.2. - Stochastic Modeling
A6.1.4. - Multiscale 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
A6.2.4. - Statistical methods
A6.2.5. - Numerical Linear Algebra
A6.3. - Computation-data interaction
A6.3.1. - Inverse problems
A6.3.2. - Data assimilation
A6.3.4. - Model reduction
A6.3.5. - Uncertainty Quantification
A6.5. - Mathematical modeling for physical sciences
A6.5.2. - Fluid mechanics
A6.5.5. - Chemistry

Other Research Topics and Application Domains:

B1. - Life sciences
B1.1. - Biology
B1.1.1. - Structural biology
B3. - Environment and planet
B3.2. - Climate and meteorology
B4. - Energy
B4.2. - Nuclear Energy Production
B4.2.1. - Fission

1. Team, Visitors, External Collaborators

Research Scientists

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

Faculty Member

Valérie Monbet [Univ Rennes, Professor, HDR]

External Collaborator

Arnaud Guyader [Univ Pierre et Marie Curie, HDR]

PhD Students

Audrey Cuillery [DCNS Group, granted by CIFRE]
Marie Morvan [Univ Rennes]
Gabriel Jouan [Univ Rennes, Scalian, granted by CIFRE]
Esso-Ridah Bleza [Univ Bretagne Sud, Janasense, granted by CIFRE]
Said Obakrim [Univ Rennes, Ifremer]

Administrative Assistant

Fabienne Cuyollaa [Inria, Administrative Assistant]

2. Overall Objectives

2.1. Overall Objectives

As the constant surge of computational power is nurturing scientists into simulating the most detailed features of reality, from complex molecular systems to climate or weather forecast, the computer simulation of physical systems is becoming reliant on highly complex stochastic dynamical models and very abundant observational data. The complexity of such models and of the associated observational data stems from intrinsic physical features, which do include high dimensionality as well as intricate temporal and spatial multi-scales. It also results in much less control over simulation uncertainty.

Within this highly challenging context, SIMSMART positions itself as a mathematical and computational probability and statistics research team, dedicated to *Monte Carlo simulation* methods. Such methods include in particular particle Monte Carlo methods for rare event simulation, data assimilation and model reduction, with application to stochastic random dynamical physical models. The main objective of SIMSMART is to disrupt this now classical field by creating deeper mathematical frameworks adapted to the management of contemporary highly sophisticated physical models.

3. Research Program

3.1. Research Program

Introduction. Computer simulation of physical systems is becoming increasingly reliant on highly complex models, as the constant surge of computational power is nurturing scientists into simulating the most detailed features of reality – from complex molecular systems to climate/weather forecast.

Yet, when modeling physical reality, bottom-up approaches are stumbling over intrinsic difficulties. First, the timescale separation between the fastest simulated microscopic features, and the macroscopic effective slow behavior becomes huge, implying that the fully detailed and direct long time simulation of many interesting systems (*e.g.* large molecular systems) are out of reasonable computational reach. Second, the chaotic dynamical behaviors of the systems at stake, coupled with such multi-scale structures, exacerbate the intricate uncertainty of outcomes, which become highly dependent on intrinsic chaos, uncontrolled modeling, as well as numerical discretization. Finally, the massive increase of observational data addresses new challenges to classical data assimilation, such as dealing with high dimensional observations and/or extremely long time series of observations.

SIMSMART Identity. Within this highly challenging applicative context, SIMSMART positions itself as a computational probability and statistics research team, with a mathematical perspective. Our approach is based on the use of *stochastic modeling* of complex physical systems, and on the use of *Monte Carlo simulation* methods, with a strong emphasis on dynamical models. The two main numerical tasks of interest to SIMSMART are the following: (i) simulating with pseudo-random number generators - a.k.a. *sampling* - dynamical models of random physical systems, (ii) sampling such random physical dynamical models given some real observations - a.k.a. *Bayesian data assimilation*. SIMSMART aims at providing an appropriate mathematical level of abstraction and generalization to a wide variety of Monte Carlo simulation algorithms in order to propose non-superficial answers to both *methodological and mathematical* challenges. The issues to be resolved include computational complexity reduction, statistical variance reduction, and uncertainty quantification.

SIMSMART's Objectives. The main objective of SIMSMART is to disrupt this now classical field of particle Monte Carlo simulation by creating deeper mathematical frameworks adapted to the challenging world of complex (*e.g.* high dimensional and/or multi-scale), and massively observed systems, as described in the beginning of this introduction.

To be more specific, we will classify SIMSMART objectives using the following four intertwined topics:

1. Objective 1: Rare events and random simulation.
2. Objective 2: High dimensional and advanced particle filtering.
3. Objective 3: Non-parametric approaches.
4. Objective 4: Model reduction and sparsity.

Rare events Objective 1 are ubiquitous in random simulation, either to accelerate the occurrence of physically relevant random slow phenomenons, or to estimate the effect of uncertain variables. Objective 1 will be mainly concerned with particle methods where *splitting* is used to enforce the occurrence of rare events.

The problem of high dimensional observations, the main topic in Objective 2, is a known bottleneck in filtering, especially in non-linear particle filtering, where linear data assimilation methods remain the state-of-the-art approaches.

The increasing size of recorded observational data and the increasing complexity of models also suggest to devote more effort into non-parametric data assimilation methods, the main issue of Objective 3.

In some contexts, for instance when one wants to compare solutions of a complex (*e.g.* high dimensional) dynamical systems depending on uncertain parameters, the construction of relevant reduced-order models becomes a key topic. This is the content of Objective 4.

With respect to volume of research activity, Objective 1, Objective 4 and the sum (Objective 2+Objective 3) are comparable.

Some new challenges in the simulation and data assimilation of random physical dynamical systems have become prominent in the last decade. A first issue (i) consists in the intertwined problems of simulating on large, macroscopic random times, and simulating *rare events*. The link between both aspects stems from the fact that many effective, large times dynamics can be approximated by sequences of rare events. A second, obvious, issue (ii) consists in managing *very abundant observational data*. A third issue (iii) consists in quantifying *uncertainty/sensitivity/variance* of outcomes with respect to models or noise. A fourth issue (iv) consists in managing *high dimensionality*, either when dealing with complex prior physical models, or with very large data sets. The related increase of complexity also requires, as a fifth issue (v), the construction of *reduced models* to speed-up comparative simulations. In a context of very abundant data, this may be replaced by a sixth issue (vi) where complexity constraints on modeling is replaced by the use of *non-parametric statistical inference*.

Hindsight suggests that all the latter challenges are related. Indeed, the contemporary digital condition, made of a massive increase in computational power and in available data, is resulting in a demand for more complex and uncertain models, for more extreme regimes, and for using inductive approaches relying on abundant data.

For simplicity, we have classified SIMSMART research into the following already mentioned four main objectives.

1. Objective 1: Rare events and random simulation, which mainly encompass item (i).
2. Objective 2: High dimension and advanced particle filtering, which encompass item (iv).
3. Objective 3: Non-parametric inference, which mainly encompass item (ii) and (vi).
4. Objective 4: Model reduction, which mainly encompasses item (vi).

Uncertainty quantification (item (iii)) in fact underlies each aspect since we are mainly interested in Monte Carlo approaches, so that uncertainty can be *modeled by an initial random variable and be incorporated in the state space of the physical model*.

4. Application Domains

4.1. Domain 1 – Computational Physics

The development of large-scale computing facilities has enabled simulations of systems at the *atomistic scale* on a daily basis. The aim of these simulations is to bridge the time and space scales between the macroscopic properties of matter and the stochastic atomistic description. Typically, such simulations are based on the ordinary differential equations of classical mechanics supplemented with a random perturbation modeling temperature, or collisions between particles.

Let us give a few examples. In bio-chemistry, such simulations are key to predict the influence of a ligand on the behavior of a protein, with applications to drug design. The computer can thus be used as a *numerical microscope* in order to access data that would be very difficult and costly to obtain experimentally. In that case, a rare event (Objective 1) is given by a macroscopic system change such as a conformation change of the protein. In nuclear safety, such simulations are key to predict the transport of neutrons in nuclear plants, with application to assessing aging of concrete. In that case, a rare event is given by a high energy neutron impacting concrete containment structures.

A typical model used in molecular dynamics simulation of open systems at given temperature is a stochastic differential equation of Langevin type. The large time behavior of such systems is typically characterized by a hopping dynamics between 'metastable' configurations, usually defined by local minima of a potential energy. In order to bridge the time and space scales between the atomistic level and the macroscopic level, specific algorithms enforcing the realization of rare events have been developed. For instance, splitting particle methods (Objective 1) have become popular within the computational physics community only within the last few years, partially as a consequence of interactions between physicists and Inria mathematicians in ASPI (parent of SIMSMART) and MATHERIALS project-teams.

4.2. Domain 2 – Meteorology

The traditional trend in data assimilation in geophysical sciences (climate, meteorology) is to use as prior information some very complex deterministic models formulated in terms of fluid dynamics and reflecting as much as possible the underlying physical phenomenon (see *e.g.*https://www.metoffice.gov.uk/research/modelling-systems/unified-model/). Weather/climate forecasting can then be recast in terms of a Bayesian filtering problem (see Objective 2) using weather observations collected *in situ*.

The main issue is therefore to perform such Bayesian estimations with very expensive infinite dimensional prior models, and observations in large dimension. The use of some linear assumption in prior models (Kalman filtering) to filter non-linear hydrodynamical phenomena is the state-of-the-art approach, and a current field of research, but is plagued with intractable instabilities.

This context motivates two research trends: (i) the introduction of non-parametric, model-free prior dynamics constructed from a large amount of past, recorded real weather data; and (ii) the development of appropriate non-linear filtering approaches (Objective 2 and Objective 3).

SIMSMART will also test its new methods on multi-source data collected in North-Atlantic paying particular attention to coastal areas (*e.g.* within the inter-Labex SEACS).

5. New Results

5.1. Objective 1 – Rare events simulation

In [2], we present a short historical perpective of the importance splitting approach to simulate and estimate rare events, with a detailed description of several variants. We then give an account of recent theoretical results on these algorithms, including a central limit theorem for Adaptive Multilevel Splitting (AMS). Considering the asymptotic variance in the latter, the choice of the importance function, called the reaction coordinate in molecular dynamics, is also discussed. Finally, we briefly mention some worthwhile applications of AMS in various domains.

Adaptive Multilevel Splitting (AMS for short) is a generic Monte Carlo method for Markov processes that simulates rare events and estimates associated probabilities. Despite its practical efficiency, there are almost no theoretical results on the convergence of this algorithm. In [1], we prove both consistency and asymptotic normality results in a general setting. This is done by associating to the original Markov process a level-indexed process, also called a stochastic wave, and by showing that AMS can then be seen as a Fleming-Viot type particle system. This being done, we can finally apply general results on Fleming-Viot particle systems that we have recently obtained. In [1] we extend the central limit theorem to the case of synchronized branchings, where re-sampling of particles is performed after any given number of particles have been killed. The result is obtained in the generic case of Fleming-Viot particle systems.

Probability measures supported on submanifolds can be sampled by adding an extra momentum variable to the state of the system, and discretizing the associated Hamiltonian dynamics with some stochastic perturbation in the extra variable. In order to avoid biases in the invariant probability measures sampled by discretizations of these stochastically perturbed Hamiltonian dynamics, a Metropolis rejection procedure can be considered. The so-obtained scheme belongs to the class of generalized Hybrid Monte Carlo (GHMC) algorithms. In [5], we show here how to generalize to GHMC a procedure suggested by Goodman, Holmes-Cerfon and Zappa for Metropolis random walks on submanifolds, where a reverse projection check is performed to enforce the reversibility of the algorithm for large timesteps and hence avoid biases in the invariant measure. We also provide a full mathematical analysis of such procedures, as well as numerical experiments demonstrating the importance of the reverse projection check on simple toy examples.

In [24], we consider Langevin processes, which are widely used in molecular simulation to compute reaction kinetics using rare event algorithms. We prove convergence in distribution in the overdamped 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.

5.2. Objective 2 – High dimensional filtering

The work presented in [10] is about solutions for 2D multitarget tracking from image observations including its application on radar data and results on simulated data. Tracking from image observations rather than from detected points is often referred to as track–before–detect (TBD). The objective is to capture targets with a low signal–to–noise ratio (SNR) which would not be detected or tracked after data thresholding. The highly nonlinear filtering equations are approximated using a particle filter implementation. This nonlinearity emerges from the observation function which relies on the radar treatment chain, involving matched filtering of a chirp signal, and beamforming achieved from a linear phased array, leading to the raw data image. Amplitudes and phases of target–returned signals are considered as temporally fluctuating, random and unknown, creating non–deterministic contributions of the targets to the signal to deal with.

5.3. Objective 3 – Non-paramatric statistics

Production forecast errors are the main hurdle to integrate variable renewable energies into electrical power systems. Regardless of the technique, these errors are inherent in the forecast exercise, although their magnitude significantly vary depending on the method and the horizon. As power systems have to balance out these errors, their dynamic and stochastic modeling is valuable for the real time operation. The study in [23] proposes a Markov Switching Auto Regressive – MS-AR – approach. After having validated its statistical relevance, this model is used to solve the problem of the optimal management of a storage associated with a wind power plant when this virtual power plant must respect a production commitment.

5.4. Objective 4 – Model Reduction

Model reduction aims at proposing efficient algorithmic procedures for the resolution (to some reasonable accuracy) of high-dimensional systems of parametric equations. This overall objective entails many different subtasks:

1) the identification of low-dimensional surrogates of the target "solution" manifold 2) The devise of efficient methodologies of resolution exploiting low-dimensional surrogates 3) The theoretical validation of the accuracy achievable by the proposed procedures

This year, we made several contributions to these subtasks. In most of our contributions, we deviated from the standard working hypothesis involving a linear subspace surrogate.

In a first group of publications, we concentrated our attention on the so-called "sparse" low-dimensional model. In this context, we have proposed several new algorithmic solutions to decrease the computational complexity associated to projection onto this low-dimensional model. These methodologies take place in the context of "screening" procedures for LASSO. We first introduced a new screening strategy, dubbed "joint screening test", which allows the rejection of a set of atoms by performing one single test, see [4]. Our approach enables to find good compromises between complexity of implementation and effectiveness of screening. Second, we proposed two new methods to decrease the computational cost inherent to the construction of the (so-called) "safe region". Our numerical experiments show that the proposed procedures lead to significant computational gains as compared to standard methodologies, see [11]. We finally showed in another work that the main concepts underlying screening procedures can be extended to different families of convex optimization problems, see [22].

Another avenue of research has been the study of the sparse surrogate in the context of "continuous" dictionaries, where the elementary signals forming the decomposition catalog are functions of some parameters taking its values in some continuously-valued domain. In this context, we contributed to the theoretical characterization of the performance of some well-known algorithmic procedure, namely "orthogonal matching pursuit" (OMP). More specifically, we proposed the first theoretical analysis of the behavior of OMP in the continuous setup, see [12], [17], [21]. We also provided a new connection between two popular low-rank approximations of continuous dictionaries, namely the "polar" and "SVD" approximations, see [9].

The tools exploited in the field of model-order reduction and sparsity have found some particular applicative field in geophysics and fluid mechanics. In [8], [7], we derived procedures based on sparse representations to localize the positions of particles in a moving fluid. In [16], [15], [14], [26], [27], we designed learning methodologies to learn the dynamical model underlying a set of observed data.

5.5. Miscellaneous

In [25], we devise methods of variance reduction for the Monte Carlo estimation of an expectation of the type $\mathbb{E}[\phi(X,Y)]$, when the distribution of X is exactly known. The key general idea is to give each individual of a sample a weight, so that the resulting weighted empirical distribution has a marginal with respect to the variable X as close as possible to its target. We prove several theoretical results on the method, identifying settings where the variance reduction is guaranteed. We perform numerical tests comparing the methods and demonstrating their efficiency.

In [6], we consider the problem of predicting a categorical variable based on groups of inputs. Some methods have already been proposed to elaborate classifi- cation rules based on groups of variables (e.g. group lasso for logistic regression). However, to our knowledge, no tree-based approach has been proposed to tackle this issue. Here, we propose the Tree Penalized Linear Discriminant Analysis algorithm (TPLDA), a new-tree based approach which constructs a classification rule based on groups of variables.

6. Bilateral Contracts and Grants with Industry

6.1. Bilateral Contracts with Industry

1. **Scalian Alyotech**, through the CIFRE PhD project of Gabriel Jouan, dedicated to weather forecast corrections.
2. **Naval Group Research**, through the CIFRE PhD project of Audrey Cuillery dedicated to Bayesian tracking.
3. **Eau du Ponant**, through the R&D project MEDISA (https://www.eauduponant.fr/fr/actualite/lancement-du-projet-de-rd-medisa) on water industry.
4. **Cooper Standard**, Machine Learning for joints design.

6.2. Bilateral Grants with Industry

1. **EURAMED** (a Euro-Mediterranean Cooperation Initiative, which aims to develop an Internet-based, multi-parametric electronic platform for optimum design of desalination plants, supplied by Renewable Energy Sources (RES). PI: E. Koutroulis (GREECE).

7. Partnerships and Cooperations

7.1. Regional Initiatives

Inter-Labex SEACS: V. Monbet, F. Le Gland, C. Herzet and Thi Tuyet Trang Chau (PhD student) are part of the *inter Labex Cominlabs-Lebesgue-Mer SEACS, http://www.seacs.cominlabs.ueb.eu/fr*, which stands for Stochastic modEl-dAta-Coupled representationS for the analysis, simulation and reconstruction of upper ocean dynamics. This project which concerns mainly Objectives 2 and 3, aims at exploring novel statistical and stochastic methods to address the emulation, reconstruction and forecast of fine-scale upper ocean dynamics.maths-computer-sea science for ocean dynamics.

CMEMS 3DA (2018-2019): C. Herzet is part of the project *CMEMS 3DA* on data assimilation of oceanographic events with non-parametric data assimilation methods. The goal of the project is to demonstrate the relevance of data-driven strategies to improve satellite derived interpolated products and especially the geostrophic surface currents. The project is made in collaboration with IMT Atlantique Brest, Ifremer and the Institue of Geosciences and Environment in Grenoble.

Action Exploratoire – Labex Cominlabs: C. Herzet is part of a project on sparse representations in continuous dictionaries. Partners: R. Gribonval (Inria Rennes PANAMA), A. Drémeau (IMT Atlantique) and P. Tandeo (IMT Atlantique).

7.2. National Initiatives

7.2.1. ANR

ANR BECOSE (2016-2020): Beyond Compressive Sensing: Sparse approximation algorithms for ill-conditioned inverse problems.

Cédric Herzet is part of the BECOSE project. 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.

ANR Melody (2020-2024): Bridging geophysics and MachinE Learning for the modeling, simulation and reconstruction of Ocean DYnamics.

Cédric Herzet is part of the MELODY project. The MELODY project aims to bridge the physical model-driven paradigm underlying ocean/atmosphere science and AI paradigms with a view to developing geophysically-sound learning-based and data-driven representations of geophysical flows accounting for their key features (e.g., chaos, extremes, high-dimensionality).

7.3. European Initiatives

7.3.1. FP7 & H2020 Projects

ERC MsMaths (2015-2019): M. Rousset is part of *ERC MSMaths* on molecular simulation (PI T. Lelièvre). With the development of large-scale computing facilities, simulations of materials at the molecular scale are now performed on a daily basis. The objective of the MSMath ERC project is to develop and study efficient algorithms to simulate such high-dimensional systems over very long, macroscopic times. ERC MsMaths especially focus on the computational issues related to 'metastable' states, that is to say specific molecular configurations that do evolve only on very large time scales. This results in a multi-timescale computational bottleneck that needs to be addressed by specific algorithms.

7.3.2. Collaborations with Major European Organizations

The agency European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) of Darmstadt. The transfer focuses on the estimation of atmospheric 3D winds from the future hyperspectral instrument (IRS on MTG-S, developed by ESA and IASI-NG on Metop-SG developed by CNES).

7.4. International Initiatives

7.4.1. Participation in Other International Programs

ECOS ARGENTINE (2018-2021): V. Monbet has obtained a funding program through the ECOS Sud - MINCyT intiative (http://www.univ-paris13.fr/cofecub-ecos/). The program involves a collaboration with the French-Argentinian Climate Institute (http://www.cima.fcen.uba.ar/UMI/), and focuses on non-parametric, analog methods, combined with data assimilation techniques to reconstruct complex meteorological dynamics (Objective 3).

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Member of the Organizing Committees

Cédric Herzet is part of the organizing committee of the iTwist'20 Workshop.

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Cédric Herzet has given:

- INSA RENNES, 5ième année de l'option Génie Mathématique, cours de Parcimonie en traitement du signal et des images, 10h de cours magistraux + responsable du module
- Ensai RENNES, Master international « Smart Data » , cours « Foundations of Smart Sensing », 9h de cours magistraux
- Ensai RENNES, Master international « Smart Data » , cours « Advanced topics in Smart Sensing » , 3h de cours magistraux
- Ensai RENNES, Master 1, « Régression pénalisée et sélection de modèles » , cours « Advanced topics in Smart Sensing », 6h de cours magistraux + 6 TPs + responsable du module

François Le Gland has given:

- 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,
- and 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 telecommunicationst.

8.2.2. Supervision

Cédric Herzet has supervised:

- Soufiane Ait Tilat, PhD, co-supervision with Frédéric Champagnat (Onera, Palaiseau)
- Milan Courcoux-Caro, PhD, co-supervision with Charles Vanwynsberghe (ENSTA Bretagne) and Alexandre Baussard (IUT de Troyes),
- Clément Elvira, postdoc, co-supervision with Rémi Gribonval (Inria Rennes) and Charles Soussen (CentraleSupélec)
- Clément Dorffer, postdoc, co-supervision with Angélique Drémeau (ENSTA Rennes)

Mathias Rousset has supervised:

- Benjamin Dufée, master 2 (with Fredéric Cérou).

François Le Gland has supervised

- Audrey Cuillery, PhD, provisional title: *Bayesian tracking from raw data*, université de Rennes 1, started in April 2016, expected defense in early 2020, funding: CIFRE grant with Naval Group, co–direction: Dann Laneuville (Naval Group, Nantes).

V. Monbet has supervised

- Gabriel Jouan, PhD, Univ Rennes, Scalian, granted by CIFRE.
- Esso-Ridah Bleza, PhD, Univ Bretagne Sud, Janasense, granted by CIFRE.
- Said Obakrim, PhD, Univ Rennes, Ifremer.

8.2.3. PhD defended

- Y. Xu (sup.: M. Rousset and P.A. Zitt): Weak over-damped asymptotic and variance reduction
- T.T.T. Chau (sup: V. Monbet): Non-parametric methodologies for reconstruction and estimation in nonlinear state-space models
- M. Morvan (sup: V. Monbet): Modèles de régression pour données fonctionnelles hétérogènes. Application à la modélisation de données de spectrométrie dans le moyen infrarouge

8.2.4. Juries

V. Monbet has been a member of the following juries:

- Anders Hildeman - On flexible random field models for spatial statistics: Spatial mixture models and deformed SPDE models, Chalmers University, Sweden.
- Shuaitao Wang - Simulation du métabolisme de la Seine par assimilation de données en continu" , Mines Paris-Tech (reviewer)
- Alban FARCHI - Localisation des méthodes d'assimilation de données d'ensemble, ENPC

François Le Gland has been a member of the following juries:

- (Reviewer) Julien Lesouple (université de Toulouse, adviser : Jean-Yves Tourneret)
- Émilien Flayac (université Paris–Saclay, Orsay, advisers : Frédéric Jean and Karim Dahia)
- Thi Tuyet Trang Chau (université de Rennes 1, advisers : Valérie Monbet and Pierre Ailliot).

8.3. Popularization

8.3.1. Interventions

Patrick Héas has made a CS workshop for middle school students based on the ideas of 'Computer Science Unplugged'.

9. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] F. Cérou, B. Delyon, A. Guyader, M. Rousset. *On the Asymptotic Normality of Adaptive Multilevel Splitting*, in "SIAM/ASA Journal on Uncertainty Quantification", 2019, vol. 7, n^{o} 1, p. 1-30, https://arxiv.org/abs/1804.08494 - 38 pages, 5 figures [*DOI :* 10.1137/18M1187477], https://hal.archives-ouvertes.fr/hal-01774856

[2] F. Cérou, A. Guyader, M. Rousset. *Adaptive Multilevel Splitting: Historical Perspective and Recent Results*, in "Chaos", April 2019, vol. 29, n^{o} 4, p. 1-32 [*DOI :* 10.1063/1.5082247], https://hal.inria.fr/hal-01936611

[3] C. Herzet, M. M. Diallo. *Performance guarantees for a variational "multi-space" decoder*, in "Advances in Computational Mathematics", 2019, p. 1-23, forthcoming, https://hal.inria.fr/hal-02394378

[4] C. Herzet, C. Dorffer, A. Drémeau. *Gather and Conquer: Region-Based Strategies to Accelerate Safe Screening Tests*, in "IEEE Transactions on Signal Processing", June 2019, vol. 67, n^{o} 12, p. 3300-3315 [*DOI :* 10.1109/TSP.2019.2914885], https://hal-ensta-bretagne.archives-ouvertes.fr/hal-02303198

[5] T. LELIÈVRE, M. ROUSSET, G. STOLTZ.*Hybrid Monte Carlo methods for sampling probability measures on submanifolds*, in "Numerische Mathematik", June 2019, vol. 143, n^{o} 2, p. 379–421, https://arxiv.org/abs/1807.02356 [*DOI :* 10.1007/s00211-019-01056-4], https://hal.archives-ouvertes.fr/hal-01832820

[6] A. POTERIE, J.-F. DUPUY, V. MONBET, L. ROUVIÈRE.*Classification tree algorithm for grouped variables*, in "Computational Statistics", 2019, p. 1613–1648, forthcoming [*DOI :* 10.1007/s00180-019-00894-Y], https://hal.archives-ouvertes.fr/hal-01623570

International Conferences with Proceedings

[7] S. AIT TILAT, F. CHAMPAGNAT, C. HERZET.*A new sparsity based particle image reconstruction approach for particle detection*, in "ISPIV 2019 - 13th International Symposium on Particle Image Velocimetry", Munich, Germany, July 2019, p. 1-10, https://hal.inria.fr/hal-02389648

[8] S. AIT TILAT, C. HERZET, F. CHAMPAGNAT.*Détection et localisation en imagerie de particules par approximation de Taylor de dictionnaires continus*, in "GRETSI 2019", Lille, France, August 2019, p. 1-4, https://hal.archives-ouvertes.fr/hal-02362086

[9] F. CHAMPAGNAT, C. HERZET.*Atom selection in continuous dictionaries : reconciling polar and SVD approximations*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech, and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, p. 5516-5520 [*DOI :* 10.1109/ICASSP.2019.8683402], https://hal.inria.fr/hal-02059703

[10] A. CUILLERY, F. LE GLAND.*Track–before-detect on radar image observation with an adaptive auxiliary particle filter*, in "FUSION 2019 - 22nd International Conference on Information Fusion", Ottawa, Canada, July 2019, https://hal.inria.fr/hal-02413324

[11] C. DORFFER, C. HERZET, A. DRÉMEAU.*Region-based relaxations to accelerate greedy approaches*, in "27th European Signal Processing Conference, EUSIPCO 2019", La Corogne, Spain, September 2019, https://hal.inria.fr/hal-02059649

[12] C. ELVIRA, R. GRIBONVAL, C. HERZET, C. SOUSSEN.*Uniform k-step recovery with CMF dictionaries*, in "SPARS 2019 - Signal Processing with Adaptive Sparse Structured Representations", Toulouse, France, July 2019, p. 1-2, https://hal.inria.fr/hal-02157561

[13] C. ELVIRA, R. GRIBONVAL, C. SOUSSEN, C. HERZET.*OMP and continuous dictionaries: Is k-step recovery possible ?*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, p. 1-5 [*DOI :* 10.1109/ICASSP.2019.8683617], https://hal.archives-ouvertes.fr/hal-02049486

[14] S. OUALA, R. FABLET, C. HERZET, B. CHAPRON, A. PASCUAL, F. COLLARD, L. GAULTIER.*Learning stochastic representations of geophysical dynamics*, in "44th International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2019", Brighton, United Kingdom, May 2019, https://hal-imt-atlantique.archives-ouvertes.fr/hal-02005403

[15] S. OUALA, R. FABLET, C. HERZET, L. DRUMETZ, B. CHAPRON, A. PASCUAL, F. COLLARD, L. GAULTIER.*Sea surface dynamics reconstruction using neural networks based kalman filter*, in "IGARSS 2019 - International Geoscience and remote Sensing Symposium", Yokohama, Japan, July 2019, p. 1-5 [*DOI :* 10.13140/RG.2.2.11461.09446], https://hal-imt-atlantique.archives-ouvertes.fr/hal-02285697

[16] S. OUALA, D. NGUYEN, L. DRUMETZ, C. HERZET, A. PASCUAL, L. GAULTIER, F. COLLARD, B. CHAPRON, R. FABLET.*Learning ocean dynamical priors from noisy data using assimilation-derived neural nets*, in "IGARSS 2019 - International Geoscience and remote Sensing Symposium", Yokohama, Japan, July 2019, p. 1-3 [*DOI :* 10.13140/RG.2.2.35788.05767], https://hal-imt-atlantique.archives-ouvertes.fr/hal-02285693

National Conferences with Proceeding

[17] C. ELVIRA, R. GRIBONVAL, C. SOUSSEN, C. HERZET.*Identification de supports en k e´tapes avec OMP pour les dictionnaires continus*, in "GRETSI 2019 - XXVIIème Colloque francophone de traitement du signal et des images", Lille, France, August 2019, p. 1-4, https://hal.inria.fr/hal-02157571

Conferences without Proceedings

[18] G. JOUAN, A. CUZOL, V. MONBET, G. MONNIER.*Weather types prediction at medium-range from ensemble forecasts*, in "9th International workshop on Climate Informatics", Paris, France, October 2019, https://hal.archives-ouvertes.fr/hal-02425230

Other Publications

[19] P. AILLIOT, M. BOUTIGNY, E. KOUTROULIS, A. MALISOVAS, V. MONBET.*Stochastic Weather Generator for the Design and Reliability Evaluation of Desalination Systems with Renewable Energy Sources*, December 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02425916

[20] F. CÉROU, A. GUYADER, M. ROUSSET.*On Synchronized Fleming-Viot Particle Systems*, November 2019, https://arxiv.org/abs/1911.05366 - 39 pages, 3 figures, https://hal.archives-ouvertes.fr/hal-02362847

[21] C. ELVIRA, R. GRIBONVAL, C. SOUSSEN, C. HERZET.*When does OMP achieve support recovery with continuous dictionaries?*, April 2019, working paper or preprint, https://hal.inria.fr/hal-02099464

[22] C. ELVIRA, C. HERZET.*Safe Squeezing for Antisparse Coding*, November 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02368134

[23] R. LE GOFF LATIMIER, E. LE BOUËDEC, V. MONBET.*Modeling of the wind power forecast errors and associated optimal storage strategy*, June 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02151567

[24] M. ROUSSET, Y. XU, P.-A. ZITT.*A Weak Overdamped Limit Theorem for Langevin Processes*, March 2019, https://arxiv.org/abs/1709.09866 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01596954

[25] M. ROUSSET, Y. XU, P.-A. ZITT.*Reducing Variance by Reweighting Samples*, March 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01925646

[26] Y. ZHEN, P. TANDEO, S. LEROUX, J. LE SOMMER, P. AILLIOT, R. FABLET, B. CHAPRON, C. HERZET, T. PENDUFF, J. VERRON, S. METREF.*Analog forecasting (AnDA) and large-ensemble ocean simulations to improve satellite-derived gridded products*, May 2019, 1, OceanPredict '19 - GODAE OceanView Symposium, Poster, https://hal-imt-atlantique.archives-ouvertes.fr/hal-02156704

[27] Y. ZHEN, P. TANDEO, S. LEROUX, J. LE SOMMER, S. METREF, P. AILLIOT, R. FABLET, B. CHAPRON, C. HERZET, T. PENDUFF, J. VERRON.*3DA: Data-Driven Data Assimilation Applications to the Lorenz system*

and simulated sea surface heights, April 2019, 1, EGU 2019 : European Geosciences Union, Poster, https://hal-imt-atlantique.archives-ouvertes.fr/hal-02108347

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

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. Team, Visitors, External Collaborators

Research Scientists

Christine Guillemot [Team leader, Inria, Senior Researcher, HDR]
Xiaoran Jiang [Inria, Starting Research Position]
Maja Krivokuca [Inria, Starting Research Position, from Mar 2019]
Thomas Maugey [Inria, Researcher]
Aline Roumy [Inria, Researcher, HDR]

Technical Staff

Pierre Allain [Inria, Engineer]
Laurent Guillo [CNRS]
Fatma Hawary [Inria, Engineer, from Mar 2019]
Hoai Nam Nguyen [Inria, Engineer]
Mira Rizkallah [Inria, Engineer]
Alexander Sagel [Inria, Engineer, from Feb 2019]

PhD Students

Pierre David [Inria]
Elian Dib [Inria]
Simon Evain [Inria]
Patrick Garus [Orange Labs]
Navid Mahmoudian Bidgoli [Inria, until Sep 2019]
Xuan Hien Pham [InterDigital]
Jinglei Shi [Inria]

Post-Doctoral Fellows

Christian Galea [Inria, until Jan 2019]
Ehsan Miandji [Inria, from Nov 2019]
Mai Quyen Pham [Inria, until Mar 2019]

Visiting Scientist

Zhaolin Xiao [Xian University, Prof., Dec. 2018 - Nov. 2019]

Administrative Assistant

Huguette Bechu [Inria]

2. Overall Objectives

2.1. Introduction

Efficient processing, i.e., analysis, storage, access and transmission of visual content, with continuously increasing data rates, in environments which are more and more mobile and distributed, remains a key challenge of the signal and image processing community. New imaging modalities, High Dynamic Range (HDR) imaging, multiview, plenoptic, light fields, 360^{o} videos, generating very large volumes of data contribute to the sustained need for efficient algorithms for a variety of processing tasks.

Building upon a strong background on signal/image/video processing and information theory, the goal of the SIROCCO team is to design mathematically founded tools and algorithms for visual data analysis, modeling, representation, coding, and processing, with for the latter area an emphasis on inverse problems related to super-resolution, view synthesis, HDR recovery from multiple exposures, denoising and inpainting. Even if 2D imaging is still within our scope, the goal is to give a particular attention to HDR imaging, light fields, and 360^{o} videos. The project-team activities are structured and organized around the following inter-dependent research axes:

- Visual data analysis
- Signal processing and learning methods for visual data representation and compression
- Algorithms for inverse problems in visual data processing
- Distributed coding for interactive communication.

While aiming at generic approaches, some of the solutions developed are applied to practical problems in partnership with industry (InterDigital, Ateme, Orange) 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 network adaptation, of advanced functionalities such as navigation, interactive streaming and high quality rendering.

2.2. Visual Data Analysis

Most visual data processing problems require a prior step of data analysis, of discovery and modeling of correlation structures. This is a pre-requisite for the design of dimensionality reduction methods, of compact representations and of fast processing techniques. These correlation structures often depend on the scene and on the acquisition system. Scene analysis and modeling from the data at hand is hence also part of our activities. To give examples, scene depth and scene flow estimation is a cornerstone of many approaches in multi-view and light field processing. The information on scene geometry helps constructing representations of reduced dimension for efficient (e.g. in interactive time) processing of new imaging modalities (e.g. light fields or 360^{o} videos).

2.3. Signal processing and learning methods for visual data representation and compression

Dimensionality reduction has been at the core of signal and image processing methods, for a number of years now, hence have obviously always been central to the research of Sirocco. These methods encompass sparse and low-rank models, random low-dimensional projections in a compressive sensing framework, and graphs as a way of representing data dependencies and defining the support for learning and applying signal de-correlating transforms. The study of these models and signal processing tools is even more compelling for designing efficient algorithms for processing the large volumes of high-dimensionality data produced by novel imaging modalities. The models need to be adapted to the data at hand through learning of dictionaries or of neural networks. In order to define and learn local low-dimensional or sparse models, it is necessay to capture and understand the underlying data geometry, e.g. with the help of manifolds and manifold clustering tools. It also requires exploiting the scene geometry with the help of disparity or depth maps, or its variations in time via coarse or dense scene flows.

2.4. Algorithms for inverse problems in visual data processing

Based on the above models, besides compression, our goal is also to develop algorithms for solving a number of inverse problems in computer vision. Our emphasis is on methods to cope with limitations of sensors (e.g. enhancing spatial, angular or temporal resolution of captured data, or noise removal), to synthesize virtual views or to reconstruct (e.g. in a compressive sensing framework) light fields from a sparse set of input views, to recover HDR visual content from multiple exposures, and to enable content editing (we focus on color transfer, re-colorization, object removal and inpainting). Note that view synthesis is a key component of multiview and light field compression. View synthesis is also needed to support user navigation and interactive streaming. It is also needed to avoid angular aliasing in some post-capture processing tasks, such as re-focusing, from a sparse light field. Learning models for the data at hand is key for solving the above problems.

2.5. Distributed coding for interactive communication

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 solutions suitable for band and energy limited networking environments. Our goal is to address theoretical issues such as the problem of modeling the correlation channel between sources, and to practical coding solutions for distributed processing and communication and for interactive streaming.

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 modeling, of signal processing, of machine learning, of coding and information theory. Some of the proposed research axes are also based on scientific foundations of computer vision (e.g. multi-view modeling 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. Data Dimensionality Reduction

Manifolds, graph-based transforms, compressive sensing

Dimensionality reduction encompasses a variety of methods for low-dimensional data embedding, such as sparse and low-rank models, random low-dimensional projections in a compressive sensing framework, and sparsifying transforms including graph-based transforms. These methods are the cornerstones of many visual data processing tasks (compression, inverse problems).

Sparse representations, *compressive sensing*, and *dictionary learning* have been shown to be powerful tools for efficient processing of visual data. The objective of *sparse representations* is to find a sparse approximation of a given input data. In theory, given a dictionary matrix $A \in \mathbb{R}^{m \times n}$, and a data $\mathbf{b} \in \mathbb{R}^m$ with $m << n$ and A is of full row 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 $\mathbf{x}$, i.e. the number of non-zero components in $\mathbf{x}$. A is known as the dictionary, its columns a_j are the atoms, they are assumed to be normalized in Euclidean norm. There exist many solutions x to $Ax = b$. The problem is to find the sparsest solution x, i.e. the one having the fewest nonzero components. In practice, one actually seeks an approximate and thus even sparser solution which satisfies $\min\{\|\mathbf{x}\|_0 \ : \ \|A\mathbf{x} - \mathbf{b}\|_p \leq \rho\}$, for some $\rho \geq 0$, characterizing an admissible reconstruction error.

The recent theory of *compressed sensing*, in the context of discrete signals, can be seen as an effective dimensionality reduction technique. The idea behind compressive sensing is that a signal can be accurately recovered from a small number of linear measurements, at a rate much smaller than what is commonly prescribed by the Shannon-Nyquist theorem, provided that it is sparse or compressible in a known basis. Compressed sensing has emerged as a powerful framework for signal acquisition and sensor design, with a number of open issues such as learning the basis in which the signal is sparse, with the help of dictionary learning methods, or the design and optimization of the sensing matrix. The problem is in particular investigated in the context of light fields acquisition, aiming at novel camera design with the goal of offering a good trade-off between spatial and angular resolution.

While most image and video processing methods have been developed for cartesian sampling grids, new imaging modalities (e.g. point clouds, light fields) call for representations on irregular supports that can be well represented by *graphs*. Reducing the dimensionality of such signals require designing novel transforms yielding compact signal representation. One example of transform is the Graph Fourier transform whose basis functions are given by the eigenvectors of the graph Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is a diagonal degree matrix whose i^{th} diagonal element is equal to the sum of the weights of all edges incident to the node i, and $\mathbf{A}$ the adjacency matrix. The eigenvectors of the Laplacian of the graph, also called Laplacian eigenbases, are analogous to the Fourier bases in the Euclidean domain and allow representing the signal residing on the graph as a linear combination of eigenfunctions akin to Fourier Analysis. This transform is particularly efficient for compacting smooth signals on the graph. The problems which therefore need to be addressed are (i) to define graph structures on which the corresponding signals are smooth for different imaging modalities and (ii) the design of transforms compacting well the signal energy with a tractable computational complexity.

3.3. Deep neural networks

Autoencoders, Neural Networks, Recurrent Neural Networks

From dictionary learning which we have investigated a lot in the past, our activity is now evolving towards deep learning techniques which we are considering for dimensionality reduction. We address the problem of unsupervised learning of transforms and prediction operators that would be optimal in terms of energy compaction, considering autoencoders and neural network architectures.

An autoencoder is a neural network with an encoder g_e, parametrized by θ, that computes a representation Y from the data X, and a decoder g_d, parametrized by ϕ, that gives a reconstruction $\widehat{X}$ of X (see Figure below). Autoencoders can be used for dimensionality reduction, compression, denoising. When it is used for compression, the representation need to be quantized, leading to a quantized representation $\widehat{Y} = Q(Y)$ (see Figure below). If an autoencoder has fully-connected layers, the architecture, and the number of parameters to be learned, depends on the image size. Hence one autoencoder has to be trained per image size, which poses problems in terms of genericity.

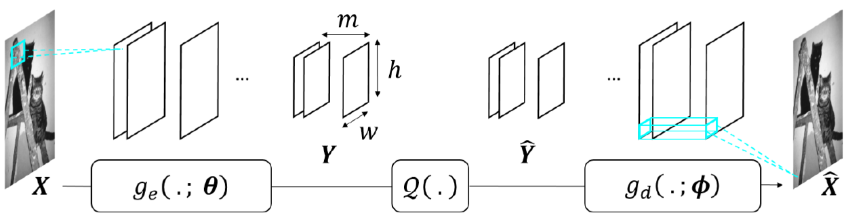


Figure 1. Illustration of an autoencoder.

To avoid this limitation, architectures without fully-connected layer and comprising instead convolutional layers and non-linear operators, forming convolutional neural networks (CNN) may be preferrable. The obtained representation is thus a set of so-called feature maps.

The other problems that we address with the help of neural networks are scene geometry and scene flow estimation, view synthesis, prediction and interpolation with various imaging modalities. The problems are posed either as supervised or unsupervised learning tasks. Our scope of investigation includes autoencoders, convolutional networks, variational autoencoders and generative adversarial networks (GAN) but also recurrent networks and in particular Long Short Term Memory (LSTM) networks. Recurrent neural networks attempting to model time or sequence dependent behaviour, by feeding back the output of a neural network layer at time t to the input of the same network layer at time t+1, have been shown to be interesting tools for temporal frame prediction. LSTMs are particular cases of recurrent networks made of cells composed of three types of neural layers called gates.

Deep neural networks have also been shown to be very promising for solving inverse problems (e.g. super-resolution, sparse recovery in a compressive sensing framework, inpainting) in image processing. Variational autoencoders, generative adversarial networks (GAN), learn, from a set of examples, the latent space or the manifold in which the images, that we search to recover, reside. The inverse problems can be re-formulated using a regularization in the latent space learned by the network. For the needs of the regularization, the learned latent space may need to verify certain properties such as preserving distances or neighborhood of the input space, or in terms of statistical modeling. GANs, trained to produce images that are plausible, are also useful tools for learning texture models, expressed via the filters of the network, that can be used for solving problems like inpainting or view synthesis.

3.4. 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 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.

[0] T. M. Cover and J. A. Thomas, Elements of Information Theory, Second Edition, July 2006.

[0] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources." IEEE Transactions on Information Theory, 19(4), pp. 471-480, July 1973.

[0] 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.

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 imaging modalities, e.g. high dynamic range (HDR) images and videos (higher bit depth, extended colorimetric space), light fields and omni-directional imaging.

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 (FVV). 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

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 and mobile access networks with different radio access technologies 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

- C. Guillemot has received the 2019 EURASIP Technical Achievement Award.

6. New Software and Platforms

6.1. M360CT

Multi-360 Calibration Toolkit

KEYWORDS: Omnidirectional camera - Calibration - FTV - 6DoF

FUNCTIONAL DESCRIPTION: Based on multiple synchronized sequences of a chessboard pattern moving in the scene, the algorithm computes the internal and external camera parameters of the different cameras under the unified spherical model. This software is composed of two executables, the first one for the individual calibration of each camera, the second one for the fusion of all the outputs of the first executable. The work has been submitted at APP with the number IDNN.FR.001.510008.S.P.2018.000.10800.

- Participants: Cédric Le Cam, Thomas Maugey and Laurent Guillo
- Contact: Thomas Maugey
- URL: http://project.inria.fr/ftv360

6.2. Context-adaptive neural network based prediction for image compression

KEYWORDS: Deep learning - Image compression - Intra prediction - Neural networks

FUNCTIONAL DESCRIPTION: This code implements (i) the learning of deep neural networks for intra-prediction of video compression and (ii) a video coder/decoder integrating the learned deep neural networks. This code allows to reproduce the results of the paper "Thierry Dumas, Aline Roumy and Christine Guillemot. Context-adaptive neural network based prediction for image compression, IEEE Transactions on Image Processing, 2019." To this end, the code implements the pre-processing of an image database (i.e. extraction of a set of pairs containing a block to be predicted and its context) to yield a training set. Then, from this learning set, the code implements the learning of one deep neural network per block size (fully-connected network at size 8x8 and smaller, and convolutional network at bigger size). The code contains the parameters of all deep networks learned from the preprocessed ILSVRC2012 training database. Then, the code contains a modified version of the HEVC video compression test model (HM 16.9), which integrates the learned prediction functions and the signalling of the intra prediction mode (for both classical HEVC intra prediction modes and the learned neural networks).

- Contact: Aline Roumy
- URL: https://github.com/thierrydumas/context_adaptive_neural_network_based_prediction

6.3. LFDE-FLEX

LFDE-FLEX: A Framework for Learning Based Depth from a Flexible Subset of Dense and Sparse Light Field Views

KEYWORDS: Light fields - Depth estimation - Deep learning

FUNCTIONAL DESCRIPTION: The code implements a learning based depth estimation framework suitable for both densely and sparsely sampled light fields. The proposed framework consists of three processing steps: initial depth estimation, fusion with occlusion handling, and refinement. The estimation can be performed from a flexible subset of input views. The fusion of initial disparity estimates, relying on two warping error measures, allows us to have an accurate estimation in occluded regions and along the contours. In contrast with methods relying on the computation of cost volumes, the proposed approach does not need any prior information on the disparity range.

- Participants: Jinglei Shi, Xiaoran Jiang and Christine Guillemot
- Contact: Jinglei Shi

6.4. EPI-SENet

EPI-based light field view extrapolation network

KEYWORDS: Light fields - Deep learning - View synthesis

FUNCTIONAL DESCRIPTION: This code implements a learning based algorithm for light field view extrapolation from axial volumes of sheared epipolar plane images (EPIs). The learned SENet network is based on tensorflow backend. The inputs of this network are multiple views in a row of a structured dense light field. The network predicts novel views in order to extend the light field baseline by a factor which can go up to 4 times the initial baseline. The code also performs digital refocusing with the original and extrapolated views. As with extended numerical aperture in classical imaging, the extrapolated light field gives refocused images with a shallower depth of field (DOF), leading to more accurate refocusing results.

- Participants: Zhaolin Xiao, Jinglei Shi, Xiaoran Jiang and Christine Guillemot
- Contact: Xiaoran Jiang

6.5. 4D-SFE

4D-SFE: 4D Scene Flow Estimator from Light Fields

KEYWORDS: Light fields - Scene Flow - Motion analysis - Depth estimation

FUNCTIONAL DESCRIPTION: This software implements a method for scene flow estimation from light fields by computing an optical flow, a disparity map and a disparity variation map for the whole light field. It takes as inputs two consecutive frames from a light field video, as well as optical flow and disparity maps estimated for each view of the light field (e.g. with a deep model like PWC-Net) and saved as .flo files. First the light field is divided into 4D clusters called superrays, then a neighboring weighted graph is built between the different clusters and finally a 4D affine model is fitted for every cluster, using the initial estimations from the optical flow and disparity estimations that are contained in the cluster and in the neighboring clusters.

- Participants: Pierre David and Christine Guillemot
- Contact: Christine Guillemot

6.6. LMVS-Net

LMVS-Net: Lightweight Neural Network for Monocular View Synthesis with Occlusion Handling

KEYWORDS: Light fields - View synthesis - Deep learning

FUNCTIONAL DESCRIPTION: This code implements the method described in A Lightweight Neural Network for Monocular View Synthesis with Occlusion Handling, allowing to perform monocular view synthesis in a stereo setting. From one input image, it computes a view laterally located, left-side or right-side, and with the required disparity range depending on user input. It is also able to retrieve a disparity map from the input image, as well as a confidence map to distinguish the occluded regions, as well as to evaluate the pixelwise accuracy of the prediction. The code was developed using Keras and TensorFlow.

- Participants: Simon Evain and Christine Guillemot
- Contact: Simon Evain

6.7. 4D-LFDE

KEYWORDS: Light fields - Depth estimation

FUNCTIONAL DESCRIPTION: This code implements a learning based solution for disparity estimation for either densely or sparsely sampled light fields from 4 corner input views. The code contains two parts "DispEstim" and "DispPropa". The DispEstim module (implemented in tensorflow) takes the 4 corner views of a light field, and estimates the disparity information for these input view positions. The DispPropa module (implemented in Matlab) then generates one disparity map per target view position by propagating corner disparity maps and by applying an occlusion-aware soft 3D reconstruction method. The final output is a .mat file which contains disparity maps for every view positions of a light field.

- Participants: Xiaoran Jiang, Christine Guillemot and Jinglei Shi
- Contact: Xiaoran Jiang

6.8. Compression of omnidirectional images on the sphere

KEYWORDS: Image compression - Omnidirectional image

FUNCTIONAL DESCRIPTION: This code implements a compression scheme of omnidirectional images. The approach operates directly on the sphere, without the need to project the data on a 2D image. More specifically, from the sphere pixelization, called healpix, the code implements the partition of the set of pixels into blocks, a block scanning order, an intra prediction between blocks, and a Graph Fourier Transform for each block residual. Finally, the image to be displayed in the viewport is generated.

- Contact: Aline Roumy

6.9. Performance evaluation of interactive video compression schemes and navigation model for omnidirectional images

KEYWORDS: Image compression - Omnidirectional image

FUNCTIONAL DESCRIPTION: This code consists of two parts. First, the code generates typical navigation paths of users viewing omnidirectional images. This generation relies on Markov modeling of the user behavior (probability to choose a first viewing direction to start the navigation, probability to choose a head motion direction, probability of continuing the head motion in the same direction, probability to stop the head motion). The second part of the code implements various criteria to evaluate the compression performance. Three criteria are computed: the distortion averaged along a set of typical navigation paths, the transmission rate related to these navigation paths, and also the storage cost of the compressed image to be able to serve any possible image request. From these 3 criteria, weighted Bjontegaard metric, and iso values are computed.

- Contact: Aline Roumy

6.10. Interactive compression for omnidirectional images and texture maps of 3D models

KEYWORDS: Image compression - Random access

FUNCTIONAL DESCRIPTION: This code implements a new image compression algorithm that allows to navigate within a static scene. To do so, the code provides access in the compressed domain to any block and therefore allows extraction of any subpart of the image. This codec implements this interactive compression for two image modalities: omnidirectional images and texture maps of 3D models. For omnidirectional images the input is a 2D equirectangular projection of the 360 image. The output is the image seen in the viewport. For 3D models, the input is a texture map and the 3D mesh. The output is also the image seen in the viewport.

The code consists of three parts: (A) an offline encoder (B) an online bit extractor and (C) a decoder. The offline encoder (i) partitions the image into blocks, (ii) optimizes the positions of the access blocks, (iii) computes a set of geometry aware predictions for each block (to cover all possible navigation paths), (iv) implements transform quantization for all blocks and their predictions, and finally (v) evaluates the encoding rates. The online bit extractor (Part B) first computes the optimal and geometry aware scanning order. Then it extracts in the bitstream, the sufficient amount of information to allow the decoding of the requested blocks. The last part of the code is the decoder (Part C). The decoder reconstructs the same scanning order as the one computed at the online bit extractor. Then, the blocks are decoded (inverse transform, geometry aware predictions, ...) and reconstructed. Finally the image in the viewport is generated.

- Contact: Aline Roumy

6.11. Platforms

6.11.1. Acquisition of multi-view sequences for Free viewpoint Television

Participants: 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 the context of the project ADT ATeP 2016-2018 (funded by Inria), such datasets were 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 have been made available to the community via a public web site. In 2019, we have published and presented our dataset at the ACM MMSys conference [28].

6.11.2. CLIM processing toolbox

Participants: Pierre Allain, Christine Guillemot, Laurent Guillo.

As part of the ERC Clim project, the EPI Sirocco is developing a light field processing toolbox. The toolbox and libraries are developed in C++ and the graphical user interface relies on Qt. As input data, this tool accepts both sparse light fields acquired with High Density Camera Arrays (HDCA) and denser light fields captured with plenoptic cameras using microlens arrays (MLA). At the time of writing, in addition to some simple functionalities, such as re-focusing, change of viewpoints, with different forms of visualization, the toolbox integrates more advanced tools for scene depth estimation from sparse and dense light fields, for super-ray segmentation and scene flow estimation, and for light field denoising and angular interpolation using anisotropic diffusion in the 4D ray space. The toolbox is now being interfaced with the C/C++ API of the tensorflow platform, in order to execute deep models developed in the team for scene depth and scene flow estimation, view synthesis, and axial super-resolution.

7. New Results

7.1. Visual Data Analysis

Scene depth, Scene flows, 3D modeling, Light-fields, 3D point clouds

7.1.1. Scene depth estimation from light fields

Participants: Christine Guillemot, Xiaoran Jiang, Jinglei Shi.

While there exist scene depth estimation methods, these methods, mostly designed for stereo content or for pairs of rectified views, do not effectively apply to new imaging modalities such as light fields. We have focused on the problem of *scene depth estimation* for every viewpoint of a dense light field, exploiting information from only a sparse set of views [24]. This problem is particularly relevant for applications such as light field reconstruction from a subset of views, for view synthesis, for 3D modeling and for compression. Unlike most existing methods, the proposed algorithm computes disparity (or equivalently depth) for every viewpoint taking into account occlusions. In addition, it preserves the continuity of the depth space and does not require prior knowledge on the depth range.

We have then proposed a learning based depth estimation framework suitable for both densely and sparsely sampled light fields. The proposed framework consists of three processing steps: initial depth estimation, efficient fusion with occlusion handling and refinement. The estimation can be performed from a flexible subset of input views. The fusion of initial disparity estimates, relying on two warping errors measures, allows us to have an accurate estimation in occluded regions and along the contours. The use of trained neural networks has the advantage of a limited computational cost at estimation time. In contrast with methods relying on the computation of cost volumes, the proposed approach does not need any prior information on the disparity range. Experimental results show that the proposed method outperforms state-of-the-art light fields depth estimation methods for a large range of baselines [15].

The training of the proposed neural networks based architecture requires having ground truth disparity (or depth) maps. Although a few synthetic datasets exist for dense light fields with ground truth depth maps, no such dataset exists for sparse light fields with large baselines. This lack of training data with ground truth depth maps is a crucial issue for supervised learning of neural networks for depth estimation. We therefore created two datasets, namely SLFD and DLFD, containing respectively sparsely sampled and densely sampled synthetic light fields. To our knowledge, SLFD is the first available dataset providing sparse light field views and their corresponding ground truth depth and disparity maps. The created datasets have been made publicly available together with the code and the trained models.

7.1.2. Scene flow estimation from light fields

Participants: Pierre David, Christine Guillemot.

We have addressed the problem of scene flow estimation from sparsely sampled video light fields. Scene flows can be seen as 3D extensions of optical flows by also giving the variation in depth along time in addition to the optical flow. Scene flows are tools needed for temporal processing of light fields. Estimating dense scene flows in light fields poses obvious problems of complexity due to the very large number of rays or pixels. This is even more difficult when the light field is sparse, i.e., with large disparities, due to the problem of occlusions. The developments in this area are also made difficult due to the lack of test data, i.e., there is no publicly available synthetic video light fields with the corresponding ground truth scene flows. In order to be able to assess the performance of the proposed method, we have therefore created synthetic video light fields from the MPI Sintel dataset. This video light field data set has been produced with the Blender software by creating new production files placing multiple cameras in the scene, controlling the disparity between the set of views.

We have then developed a local 4D affine model to represent scene flows, taking into account light field epipolar geometry. The model parameters are estimated per cluster in the 4D ray space. We have first developed a sparse to dense estimation method that avoids the difficulty of computing matches in occluded areas [18], which we have further extended by developing a dense scene flow estimation method from light fields. The local 4D affine parameters are in this case derived by fitting the model on initial motion and disparity estimates obtained by using 2D dense optical flow estimation techniques.

We have shown that the model is very effective for estimating scene flows from 2D optical flows (see Fig.2). The model regularizes the optical flows and disparity maps, and interpolates disparity variation values in occluded regions. The proposed model allows us to benefit from deep learning-based 2D optical flow estimation methods while ensuring scene flow geometry consistency in the 4 dimensions of the light field.

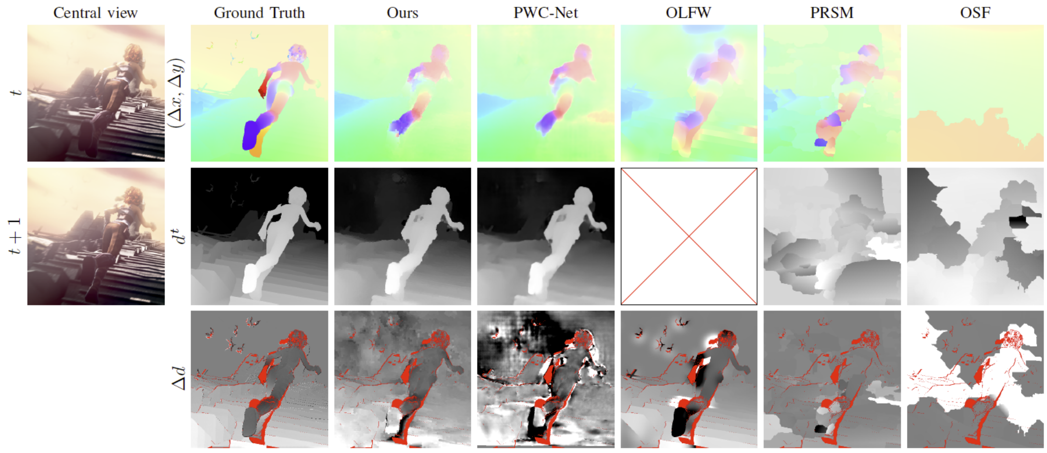


Figure 2. Visual comparison of our method with respect to reference methods (PWC-Net: deep learning method for optical flow estimation; oriented light field window (OLFW), Piece-wise Rigid Scene Model (PRSM), Object Scene Flow (OSF)). First row: optical flows; Second row: disparity maps; Third row: disparity variations. The red pixels are the occlusion mask where there is no ground truth disparity variation available.

7.1.3. Depth estimation at the decoder in the MPEG-I standard

Participants: Patrick Garus, Christine Guillemot, Thomas Maugey.

This study, in collaboration with Orange labs., addresses several downsides of the system under development in MPEG-I for coding and transmission of immersive media. We study a solution, which enables Depth-Image-Based Rendering for immersive video applications, while lifting the requirement of transmitting depth information. Instead, we estimate the depth information on the client-side from the transmitted views. We have observed that doing this leads to a significant rate saving (37.3% in average). Preserving perceptual quality in terms of MS-SSIM of synthesized views, it yields to 24.6% rate reduction for the same quality of reconstructed views after residue transmission under the MPEG-I common test conditions. Simultaneously, the required pixel rate, i.e. the number of pixels processed per second by the decoder, is reduced by 50% for any test sequence [22].

7.1.4. Spherical feature extraction for 360 light field reconstruction from omni-directional fish-eye camera captures

Participants: Christine Guillemot, Fatma Hawary, Thomas Maugey.

With the increasing interest in wide-angle or 360° scene captures, the extraction of descriptors well suited to the geometry of this content is a key problem for a variety of processing tasks. Algorithms designed for feature extraction in 2D images are hardly applicable to 360° images or videos as they do not well take into account their specific spherical geometry. To cope with this difficulty, it is quite common to perform an equirectangular projection of the spherical content, and to compute spherical features on projected and stitched content. However, this process introduces geometrical distortions with implications on the accuracy of applications such as angle estimation, depth calculation and 3D scene reconstruction. We adapt a spherical feature descriptor to the geometry of fish-eye cameras that avoids equirectangular projection. The captured image is directly mapped onto a spherical model of the 360° camera. In order to evaluate the interest of the proposed fish-eye adapted descriptor, we consider the angular coordinates of feature points on the sphere. We assess the stability of the corresponding angles when capturing the scene by a moving fish-eye camera. Experimental results show that the proposed fish-eye adapted descriptor allows a more stable angle estimation, hence a more robust feature detection, compared to spherical features on projected and stitched contents.

7.2. Signal processing and learning methods for visual data representation and compression

Sparse representation, data dimensionality reduction, compression, scalability, rate-distortion theory

7.2.1. *Single sensor light field acquisition using coded masks*

Participants: Christine Guillemot, Ehsan Miandji, Hoai Nam Nguyen.

We developed a simple variational approach for reconstructing color light fields in the compressed sensing framework with very low sampling ratio, using both coded masks and color filter arrays (CFA). A coded mask is placed in front of the camera sensor to optically modulate incoming rays, while a color filter array is assumed to be implemented at the sensor level to compress color information. Hence, the light field coded projections, operated by a combination of the coded mask and the CFA, measure incomplete color samples with a three times lower sampling ratio than reference methods that assume full color (channel-by-channel) acquisition. We then derived adaptive algorithms to directly reconstruct the light field from raw sensor measurements by minimizing a convex energy composed of two terms. The first one is the data fidelity term which takes into account the use of CFAs in the imaging model, and the second one is a regularization term which favors the sparse representation of light fields in a specific transform domain. Experimental results show that the proposed approach produces a better reconstruction both in terms of visual quality and quantitative performance when compared to reference reconstruction methods that implicitly assume prior color interpolation of coded projections.

We then pursued this study by developing a unifying image formation model that abstracts the architecture of most existing compressive-sensing light-field cameras, equipped with single lens and coded masks, as an equivalent multi-mask camera. It allows to compare different designs with a number of criteria: compression rate, light efficiency, measurement incoherence, as well as acquisition quality. Moreover, the underlying multi-mask camera can be flexibly adapted for various applications, such as single and multiple acquisitions, spatial super-resolution, parallax reconstruction, and color restoration. We also derived a generic variational algorithm solving all these concrete problems by considering appropriate sampling operators.

7.2.2. *3D point cloud processing and plenoptic point cloud compression*

Participants: Christian Galea, Christine Guillemot, Maja Krivokuca.

Light fields, by capturing light rays emitted by a 3D scene along different orientations, give a very rich description of the scene enabling a variety of computer vision applications. The recorded 4D light field gives in particular information about the parallax and depth of the scene. The estimated depth can then be used to construct 3D models of the scene, e.g. in the form of a 3D point cloud, The constructed 3D point clouds, however, generally contain distortions and artefacts primarily caused by inaccuracies in the depth maps. We have developed a method for noise removal in 3D point clouds constructed from light fields [21]. While existing methods discard outliers, the proposed approach instead attempts to correct the positions of points, and thus reduce noise without removing any points, by exploiting the consistency among views in a light-field. The proposed 3D point cloud construction and denoising method exploits uncertainty measures on depth values.

Beyond classical 3D point clouds, plenoptic point clouds can be seen as natural extensions of 3D point clouds to Surface Light Fields (SLF). While the concept of surface light field (SLF) has been introduced as a function that assigns a color to each ray originating on a surface, plenoptic point clouds represent in each voxel illumination and color seen from different camera viewpoints. In other words, instead of each point being associated with a single colour value, there can be multiple values to represent the colour at that point as perceived from different viewpoints. This concept aims at combining the best of light fields and computer graphics modeling, for photo-realistic rendering from arbitrary points of view. However, this representation leads to color maps per voxel, hence to large volumes of data. We have addressed the problem of efficient compression of this data based on the Region-Adaptive Hierarchical Transform (RAHT) method in which we have introduced clustering and speculat/diffuse components separation showing better adapted plenoptic point cloud color maps transforms.

7.2.3. Low-rank models and representations for light fields

Participants: Elian Dib, Christine Guillemot, Xiaoran Jiang.

We have addressed the problem of light field dimensionality reduction. We have introduced a local low-rank approximation method using a parameteric disparity model. The local support of the approximation is defined by super-rays. Superrays can be seen as a set of super-pixels that are coherent across all light field views. The light field low-rank assumption depends on how much the views are correlated, i.e. on how well they can be aligned by disparity compensation. We have therefore introduced a disparity estimation method using a low-rank prior. We have considered a parametric model describing the local variations of disparity within each super-ray, and alternatively search for the best parameters of the disparity model and of the low-rank approximation. We have assessed the proposed disparity parametric model, by considering an affine disparity model. We have shown that using the proposed disparity parametric model and estimation algorithm gives an alignment of superpixels across views that favours the low-rank approximation compared with using disparity estimated with classical computer vision methods. The low-rank matrix approximation is then computed on the disparity compensated super-rays using a singular value decomposition (SVD). A coding algorithm has been developed for the different components of the proposed disparity-compensated low-rank approximation [20].

We have also, in collaboration with Trinity College Dublin, introduced a new Light Field representation for efficient Light Field processing and rendering called Fourier Disparity Layers (FDL) [12]. The proposed FDL representation samples the Light Field in the depth (or equivalently the disparity) dimension by decomposing the scene as a discrete sum of layers. The layers can be constructed from various types of Light Field inputs including a set of sub-aperture images, a focal stack, or even a combination of both. From our derivations in the Fourier domain, the layers are simply obtained by a regularized least square regression performed independently at each spatial frequency, which is efficiently parallelized in a GPU implementation. Our model is also used to derive a gradient descent based calibration step that estimates the input view positions and an optimal set of disparity values required for the layer construction. Once the layers are known, they can be simply shifted and filtered to produce different viewpoints of the scene while controlling the focus and simulating a camera aperture of arbitrary shape and size. A direct implementation in the Fourier domain allows real time Light Field rendering. Finally, direct applications such as view interpolation or extrapolation and denoising have also been evaluated [12]. The use of this representation for view synthesis based compression has also been assessed in [19].

7.2.4. Graph-based transforms and prediction for light fields

Participants: Christine Guillemot, Thomas Maugey, Mira Rizkallah.

We have investigated Graph-based transforms for low dimensional embedding of light field data. Both non separable and separable transforms have been considered. The low-dimensional embedding can be learned with a few eigen vectors of the graph Laplacian. However, the dimension of the data (e.g. light fields) has obvious implications on the storage footprint of the Laplacian matrix and on the eigenvectors computation complexity, making graph-based non separable transforms impractical for such data. To cope with this difficulty, we have developed local super-rays based non separable and separable (spatial followed by angular) weighted and unweighted transforms to jointly capture light fields correlation spatially and across views [14]. Despite the local support of limited size defined by the super-rays, the Laplacian matrix of the non separable graph remains of high dimension and its diagonalization to compute the transform eigen vectors remains computationally expensive. To solve this problem, we have then performed the local spatio-angular transform in a separable manner.

Separable transforms on super-rays allow us to significantly decrease the eigenvector computation complexity. However, the basis functions of the spatial graph transforms to be applied on the super-ray pixels of each view are often not compatible. We have indeed shown that when the shape of corresponding super-pixels in the different views is not isometric, the basis functions of the spatial transforms are not coherent, resulting in decreased correlation between spatial transform coefficients, hence in a loss of performance of the angular transform, compared to the non-separable case. We have therefore developed a graph construction optimization procedure which seeks to find the eigen-vectors which align the best with those of a reference one while still approximately diagonalizing their respective Laplacians [14]. The proposed optimization method aims at preserving angular correlation even when the shapes of the super-pixels are not isometric. Experimental results show the benefit of the approach in terms of energy compaction. A coding scheme has also been developed to assess the rate-distortion perfomances of the proposed transforms

The use of local transforms with limited supports is a way to cope with the computational difficulty. Unfortunately, the locality of the support may not allow us to fully exploit long term signal dependencies present in both the spatial and angular dimensions in the case of light fields. We have therefore introduced sampling and prediction schemes, based on graph sampling theory, with local graph-based transforms enabling to efficiently compact the signal energy and exploit dependencies beyond the local graph support [31], [13]. The proposed approach has been shown to be very efficient in the context of spatio-angular transforms for quasi-lossless compression of light fields.

7.2.5. Intra-coding of 360-degree images on the sphere

Participants: Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy.

Omni-directional images are characterized by their high resolution (usually 8K) and therefore require high compression efficiency. Existing methods project the spherical content onto one or multiple planes and process the mapped content with classical 2D video coding algorithms. However, this projection induces sub-optimality. Indeed, after projection, the statistical properties of the pixels are modified, the connectivity between neighboring pixels on the sphere might be lost, and finally, the sampling is not uniform. Therefore, we propose to process uniformly distributed pixels directly on the sphere to achieve high compression efficiency. In particular, a scanning order and a prediction scheme are proposed to exploit, directly on the sphere, the statistical dependencies between the pixels. A Graph Fourier Transform is also applied to exploit local dependencies while taking into account the 3D geometry. Experimental results demonstrate that the proposed method provides up to 5.6% bitrate reduction and on average around 2% bitrate reduction over state-of-the-art methods. This work has led to a publication in the PCS conference 2019 [26].

7.3. Algorithms for inverse problems in visual data processing

Inpainting, view synthesis, super-resolution

7.3.1. View synthesis in light fields and stereo set-ups

Participants: Simon Evain, Christine Guillemot, Xiaoran Jiang, Jinglei Shi.

We have developed a learning-based framework for light field view synthesis from a subset of input views. Building upon a light-weight optical flow estimation network to obtain depth maps, our method employs two reconstruction modules in pixel and feature domains respectively. For the pixel-wise reconstruction, occlusions are explicitly handled by a disparity-dependent interpolation filter, whereas inpainting on disoccluded areas is learned by convolutional layers. Due to disparity inconsistencies, the pixel-based reconstruction may lead to blurriness in highly textured areas as well as on object contours. On the contrary, the feature-based reconstruction performs well on high frequencies, making the reconstruction in the two domains complementary. End-to-end learning is finally performed including a fusion module merging pixel and feature-based reconstructions. Experimental results show that our method achieves state-of-the-art performance on both synthetic and real-world datasets, moreover, it is even able to extend light fields baseline by extrapolating high quality views without additional training.

We have also designed a very lightweight neural network architecture, trained on stereo data pairs, which performs view synthesis from one single image [7]. With the growing success of multi-view formats, this problem is indeed increasingly relevant. The network returns a prediction built from disparity estimation, which fills in wrongly predicted regions using a occlusion handling technique. To do so, during training, the network learns to estimate the left-right consistency structural constraint on the pair of stereo input images, to be able to replicate it at test time from one single image. The method is built upon the idea of blending two predictions: a prediction based on disparity estimation, and a prediction based on direct minimization in occluded regions. The network is also able to identify these occluded areas at training and at test time by checking the pixelwise left-right consistency of the produced disparity maps. At test time, the approach can thus generate a left-side and a right-side view from one input image, as well as a depth map and a pixelwise confidence measure in the prediction. The work outperforms visually and metric-wise state-of-the-art approaches on the challenging KITTI dataset, all while reducing by a very significant order of magnitude (5 or 10 times) the required number of parameters (6.5 M).

7.3.2. Inverse problems in light field imaging with 4D anisotropic diffusion and neural networks

Participants: Pierre Allain, Christine Guillemot, Laurent Guillo.

We have addressed inverse problems in ligt field imaging by following two methodological directions. We first introduced a 4D anisotropic diffusion framework based on PDEs [4]. The proposed regularization method operated in the 4D ray space and, unlike the methods operating on epipolar plane images, does not require prior estimation of disparity maps. The method performs a PDE-based diffusion with anisotropy steered by a tensor field based on local structures in the 4D ray space that we extract using a 4D tensor structure. To enhance coherent structures, the smoothing along directions, surfaces, or volumes in the 4D ray space is performed along the eigenvectors directions. Although anisotropic diffusion is well understood for 2D imaging, its interpretation and understanding in the 4D space is far from being straightforward. We have analysed the behaviour of the diffusion process on a light field toy example, i.e. a tesseract (a 4D cube). This simple light field example allows an in-depth analysis of how each eigenvector influences the diffusion process. The proposed ray space regularizer is a tool that has enabled us to tackle a variety of inverse problems (denoising, angular and spatial interpolation, regularization for enhancing disparity estimation as well as inpainting) in the ray space.

In collaboration with the university of Malta (Pr. Reuben Farrugia), we have explored the benefit of low-rank priors in light field super-resolution with deep neural networks. This led us to design a learning-based spatial light field super-resolution method that allows the restoration of the entire light field with consistency across all sub-aperture images [8]. The algorithm first uses optical flows to align the light field views and then reduces its angular dimension using low-rank approximation. We then consider the linearly independent columns of the resulting low-rank model as an embedding, which is restored using a deep convolutional neural network. The super-resolved embedding is then used to reconstruct the remaining sub-aperture images. The original disparities are restored using inverse warping where missing pixels are approximated using a novel light field inpainting algorithm. We pursued this study by designing an approach that, thanks to a low-rank approximation model, can leverage models learned for 2D image super-resolution [9]. This approach avoids the need for a large amount of light field training data which is, unlike 2D images, not available. It also allows us to reduce the dimension, hence the number of parameters, of the network to be learned.

7.3.3. Neural networks for axial light field super-resolution

Participants: Christine Guillemot, Zhaolin Xiao.

Axial light field resolution refers to the ability to distinguish features at different depths by refocusing. The axial refocusing precision corresponds to the minimum distance in the axial direction between two distinguishable refocusing planes. High refocusing precision can be essential for some light field applications like microscopy. We first introduced a refocusing precision model based on a geometrical analysis of the flow of rays within the virtual camera. The model establishes the relationship between the feature distinguishability by refocusing and different camera settings. We have then developed a learning-based method to extrapolate novel views from axial volumes of sheared epipolar plane images (EPIs (see an example of extrapolated views in Fig.3)). As extended numerical aperture (NA) in classical imaging, the extrapolated light field gives re-focused images with a shallower depth of field (DOF), leading to more accurate refocusing results. Most importantly, the proposed approach does not need accurate depth estimation. Experimental results with both synthetic and real light fields, including with microscopic data, demonstrate that our approach can effectively enhance the light field axial refocusing precision.

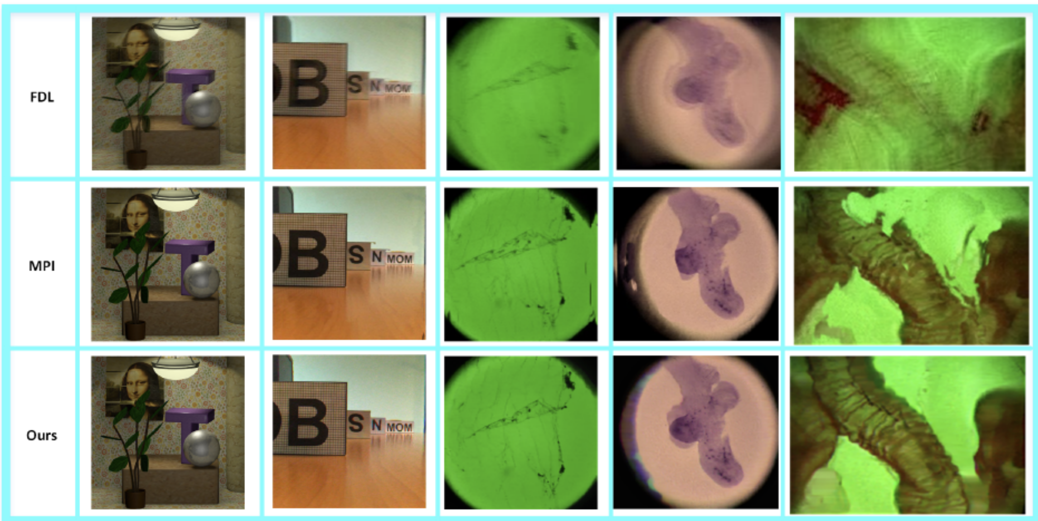


Figure 3. Extrapolation results with a 4X larger baseline, in comparison with reference methods using multiple plane images (MPI) and Fourier disparity layers (FDL).

7.3.4. Neural networks for inverse problems in 2D imaging

Participants: Christine Guillemot, Aline Roumy, Alexander Sagel.

The Deep Image Prior has been recently introduced to solve inverse problems in image processing with no need for training data other than the image itself. However, the original training algorithm of the Deep Image Prior constrains the reconstructed image to be on a manifold described by a convolutional neural network. For some problems, this neglects prior knowledge and can render certain regularizers ineffective. We have developed an alternative approach that relaxes this constraint and fully exploits all prior knowledge. We have evaluated our algorithm on the problem of reconstructing a high-resolution image from a downsampled version and observed a significant improvement over the original Deep Image Prior algorithm.

7.4. Distributed coding for interactive communication

Information theory, stochastic modeling, 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 compression scheme for interactive media

Participants: Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy.

We propose a new interactive compression scheme for omnidirectional images and 3D model. This requires two characteristics: efficient compression of data, to lower the storage cost, and random access ability to extract part of the compressed stream requested by the user (for reducing the transmission rate). For efficient compression, data needs to be predicted by a series of references that have been pre-defined and compressed. This contrasts with the spirit of random accessibility. We propose a solution for this problem based on incremental codes implemented by rate adaptive channel codes. This scheme encodes the image while adapting to any user request and leads to an efficient coding that is flexible in extracting data depending on the available information at the decoder. Therefore, only the information which is needed to be displayed at the user's side is transmitted during the user's request as if the request was already known at the encoder (see Fig. 4). The experimental results demonstrate that our coder obtains a better transmission rate than the state-of-the-art tile-based methods at a small cost in storage. Moreover, the transmission cost grows gradually with the size of the request and avoids a staircase effect, which shows the perfect suitability of our coder for interactive transmission. This work has led to a journal submission and several conference publications. In [25], we have proposed a new framework for evaluating the compression performance of interactive schemes. Indeed, interactive compression schemes can be characterized by tree criteria: the storage cost, the transmission rate and distortion. This contrasts with classical compression scheme, where only transmission rate and distortion are used. 3D-performance evaluation criteria are proposed. In [29], we have proposed to use the geometry to efficiently compress the 3D mesh texture. An interactive coding extension has been presented in [27].

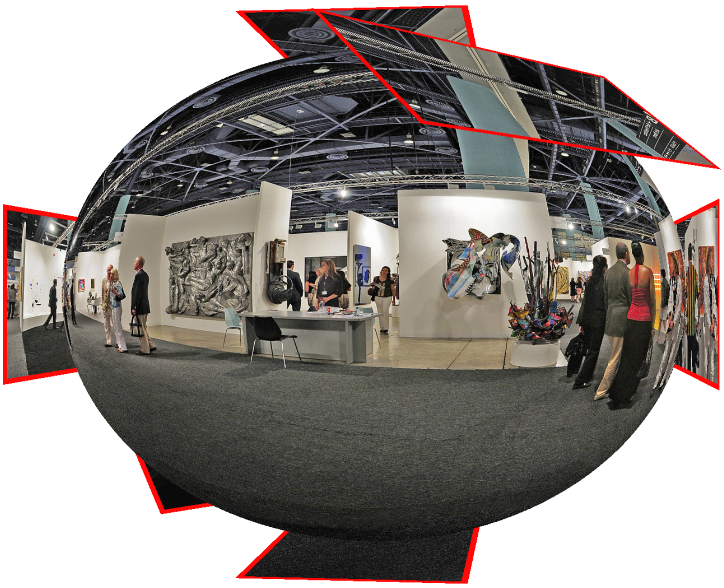


Figure 4. A spherical image and several viewports corresponding to different user's requests.

7.4.2. Reference source positioning for interactive compression

Participants: Thomas Maugey, Mai Quyen Pham, Aline Roumy.

Large databases containing many HD videos or records from sensors over long time intervals, have to be efficiently compressed, to reduce their size. The compression has also to allow efficient access to random parts of the databases upon request from the users. Efficient compression is usually achieved with prediction between data points. However, this creates dependencies between the compressed representations, which is contrary to the idea of random access. Prediction methods rely in particular on reference data points, used to predict other data points, and the placement of these references balances compression efficiency and random access. Existing solutions to position the references use ad hoc methods. We study this joint problem of compression efficiency and random access. We introduce the storage cost as a measure of the compression efficiency and the transmission cost for the random access ability. We show that the reference placement problem that trades off storage with transmission cost is an integer linear programming problem, that can be solved by standard optimizer. Moreover, we show that the classical periodic placement of the references is only optimal in a very restrictive case: namely, when the encoding costs of each data point are equal and when requests of successive data points are made.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. CIFRE contract with InterDigital on neural networks for video compression

Participants: Xuan Hien Pham, Christine Guillemot.

- Title : Neural networks for video compression
- Partners : InterDigital (Ph. Bordes, F. Galpin), Inria-Rennes.
- Funding : InterDigital, ANRT.
- Period : Jan.2019-Oct.2021.

The goal of this Cifre contract is to first investigate novel optical flow estimation methods using deep neural networks. Based on the optical flow methods, the next step will be to design temporal prediction schemes based on convolutional neural networks (CNN) for video compression. The methods will be assessed in the context of the VVC (Versatile Video Coding) standard.

8.1.2. CIFRE contract with Orange labs. on compression of immersive content

Participants: Patrick Garus, Christine Guillemot, Thomas Maugey.

- Title : Compression of immersive content
- Research axis : 7.1.3
- Partners : Orange labs. (J. Jung), Inria-Rennes.
- Funding : InterDigital, ANRT.
- Period : Jan.2019-Dec.2021.

The goal of this Cifre contract is to develop novel compression methods for 6 DoF immersive video content. This implies investigating depth estimation and view synthesis methods that would be robust to quantization noise. This also implies developing the corresponding coding mode decisions based on rate-distortion criteria.

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 : 7.4.1
- Partners : Inria-Rennes (Sirocco team and i4S team); LabSTICC, IMT Atlantique, Signal & Communications Department; External partners: L2S, CentraleSupelec, Univ. Paris Sud; EPFL, Signal Processing Laboratory (LTS4).
- Funding : Labex CominLabs.
- Period : Oct. 2016 - Dec. 2020.

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 are being 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 Allain, Pierre David, Elian Dib, Simon Evain, Christian Galea, Christine Guillemot, Laurent Guillo, Fatma Hawary, Xiaoran Jiang, Maja Krivokuca, Ehsan Miandji, Hoai Nam Nguyen, Mira Rizkallah, Alexander Sagel, Jinglei Shi.

- Title : Computational Light field Imaging.
- Research axis : 7.1.1, 7.1.2, 7.1.4, 7.2.1, 7.2.3, 7.2.4, 7.2.2, 7.3.1, 7.3.2, 7.3.3, 7.3.4
- Partners : Inria-Rennes
- Funding : European Research Council (ERC) advanced grant
- Period : Sept. 2016 - Aug. 2021.

All imaging systems, when capturing a view, record different combinations of light rays emitted by the environment. In a conventional camera, each sensor element sums all the light rays emitted by one point over the lens aperture. Light field cameras instead measure the light along each ray reaching the camera sensors and not only the sum of rays striking each point in the image. In one single exposure, they capture the geometric distribution of light passing through the lens. This process can be seen as sampling the plenoptic function that describes the intensity of the light rays interacting with the scene and received by an observer at every point in space, along any direction of gaze, for all times and every wavelength.

The recorded flow of rays (the light field) is in the form of high-dimensional data (4D or 5D for static and dynamic light fields). The 4D/5D light field yields a very rich description of the scene enabling advanced creation of novel images from a single capture, e.g. for computational photography by simulating a capture with a different focus and a different depth of field, by simulating lenses with different apertures, by creating images with different artistic intents. It also enables advanced scene analysis with depth and scene flow estimation and 3D modeling. The goal of the ERC-CLIM project is to develop algorithms for the entire static and video light fields processing chain. The planned research includes the development of:

- novel low-rank or graph-based models for dimensionality reduction and compression
- deep learning methods for scene analysis (e.g. scene depth and scene flow estimation)
- learning methods for solving a range of inverse problems: denoising, super-resolution, axial super-resolution, view synthesis.

9.3. International Initiatives

9.3.1. Inria International Labs

EPFL-Inria: Associate Team involved in the International Lab: Graph-based Omnidirectional video Processing (GOP)

- Participant: Thomas Maugey
- International Partner (Institution - Laboratory - Researcher): Ecole Polytechnique Fédérale de Lausanne (Switzerland) - LTS4 - Pascal Frossard
- period: 2017-2019

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 led 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. During this project, we study how graphs can be built for defining a suitable structure on one or several 360 videos and then used for compression.

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 started 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.
- Mikael Le Pendu and Prof. Aljosa Smolic from Trinity College Dublin on HDR light field recovery from multiple exposures.
- Pascal Frossard, Prof. at EPFL, in the context of the Comin Lab/Intercom project and in the context of the EPFL-Inria associated team.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

9.4.1.1. Internships

- Zhaolin Xiao, Prof. Xian University, Dec. 2018-Nov. 2019.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

10.1.1.1. General Chair, Scientific Chair

- C. Guillemot has organized, as chair, the ERC-CLIM workshop, May 2019.

10.1.2. Scientific Events: Selection

10.1.2.1. Member of the Conference Program Committees

- C. Guillemot has been member of the international steering committee of the Picture Coding Symposium (PCS), 2019.
- C. Guillemot has been member of the technical programme committee of IEEE-ICASSP 2019.
- C. Guillemot has been a member of the technical program committee of the CVPR 2019 workshop on New Trends in Image Restoration and Enhancement (NTIRE), 2019.
- A. Roumy has been a member of the technical program committee of the CVPR 2019 workshop on New Trends in Image Restoration and Enhancement (NTIRE).
- A. Roumy has been a member of the technical program committee of the IEEE International Conference on Communications (ICC) 2019, Workshop on Machine Learning in Wireless Communications (ML4COM).
- A. Roumy has been a member of the technical program committee of the IEEE Wireless Communications and Networking Conference 2019 (IEEE WCNC).
- A. Roumy has been a member of the technical program committee of the International conference on Telecommunications (ICT) 2019.
- A. Roumy has been a member of the technical program committee of the Gretsi 2019 conference, organized by the National Research group in Image and Signal Processing.

10.1.3. Journal

10.1.3.1. 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.
- T. Maugey is associate editor of the EURASIP Journal on Advances in Signal Processing (since Dec 2019).
- A. Roumy is associate editor of the Springer Annals of Telecommunications.
- A. Roumy is associate editor of the IEEE Trans. on Image Processing.

10.1.4. Invited Talks

- C. Guillemot gave a keynote talk on Light field image processing, at the Digital Optical Technologies International Conference, Munich, June 24-27, 2019.

- Thomas Maugey gave a seminar at IRISA's days on Art, Culture and Heritage, Rennes, France, "Data acquisition and compression for user immersion in a virtual scene" (Jan. 2019).
- Thomas Maugey gave a seminar at Orange Labs. PhD days, Rennes, France, "Light Field Acquisition and Depth estimation" (Mar. 2019).
- Thomas Maugey gave a seminar at GdR-Isis, Rennes, France, "Data acquisition and compression for user immersion in a 3D scene" (Mar. 2019).
- A. Roumy gave a talk on "Learning transforms for image compression", workshop on Learning for unstructured data, Univ. Paris 13, Nov. 2019.

10.1.5. Leadership within the Scientific Community

- C. Guillemot is member of the IEEE Signal Processing Society Nominations and Appointments Committee for a two-year term.
- A. Roumy is member of the IEEE IVMSP technical committee.
- A. Roumy is a Local Liaison Officer for the European Association for Signal Processing (EURASIP).
- A. Roumy is a member of the Executive board of the National Research group in Image and Signal Processing (GRETSI).

10.1.6. Scientific Expertise

- C. Guillemot has served as expert in a research proposal selection committee of the Irish Research Council, Feb. 2019.
- C. Guillemot is responsible of the theme « compression et protection des données images » for the realization of the encyclopedia SCIENCES of the publisher ISTE/Wiley (2019-2021).

10.1.7. Research Administration

- C. Guillemot is member of the "bureau du Comité des Projets".
- C. Guillemot has served as a member of the selection committee of a Prof. at the National University of Ireland (NUI), Galway, Ireland, Feb. 2019.
- C. Guillemot has served as a member of the selection committee of a Prof. at the Technical Univ. of Denmark, May 2019.
- A. Roumy is a member of Inria evaluation committee.
- A. Roumy has served as a member of the selection committee of an Assistant Prof. at University Paris 13 (May 2019).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

- Master: C. Guillemot, Image and video compression, 10 hours (2018-2019), and advanced video processing, 4 hours (2018-2019), M2 SISEA, Univ. of Rennes 1, France.
- Undergraduate: L. Guillo, course of 35 hours on Functional immutable programming, 1st year of Mathematiques, Informatique, Electronique, mathematique-Economie" (MIEE), Univ. of Rennes 1
- Master: T. Maugey, course on 3D models in a module on advanced video, 4 hours (2018-2019), 12 hours (2019-2020), M2 SISEA, Univ. of Rennes 1, France.
- Master: T. Maugey, course on Image compression, 10 hours (2019-2020), M2 SISEA, Univ. of Rennes 1, France.
- Master: T. Maugey, course on Representation, editing and perception of digital images, 12 hours, M2 SIF, 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.
- Master: A. Roumy, Foundations of smart sensing, 18 hours, ENSAI, Master of Science in Statistics for Smart Data, France.
- Master: A. Roumy, Information theory, 15 hours, University Rennes 1, SIF master, France.

10.2.2. Juries

- C. Guillemot has been reviewer of the PhDs of:
  - B. Ray, UNiv. Poitiers, March 2019
  - T. Trinh Le, Univ. Paris Saclay, June 2019
  - Z. Chen, Chinese Univ. of Hong Kong, June 2019
  - D. Tsai, Queensland Univ. of Technology (QUT), Nov. 2019
  - M. De Pinheiro Carvalho, Univ. Paris Saclay, Nov. 2019
- C. Guillemot has been member of the PhD committees of:
  - M. Bichon, Univ. Rennes 1, March 2019
  - C.A. Noury, Univ. Clermond Ferrand, Nov. 2019
  - S. Manandhar, Univ. Rennes 1, Nov. 2019
- Aline Roumy has been reviewer of the PhDs of:
  - Shuo Zheng, Univ. Paris Saclay, prepared at Telecom ParisTech, Feb. 2019
  - Mohsen Abdoli , Univ. Paris Saclay, prepared at Univ. Paris Sud, June 2019
  - David Kibloff , Univ. Lyon, prepared at Insa Lyon, Sept. 2019
- Aline Roumy has been member of the PhD committees of:
  - Yigit Ugur, Univ. Paris Est, Nov. 2019
  - Fangping Ye, IMT Atlantique, Brest, Dec. 2019

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

- C. Guillemot is responsible of the theme « compression et protection des données images » for the realization of the encyclopedia SCIENCES of the publisher ISTE/Wiley (2019-2021).

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] F. HAWARY.*Light Field Image Compression and Compressive Acquisition*, Université de Rennes 1, France ; Inria, May 2019, https://hal.archives-ouvertes.fr/tel-02378409

[2] N. MAHMOUDIAN BIDGOLI.*Compression for interactive communication of visual contents*, Université de Rennes 1, November 2019, https://hal.inria.fr/tel-02410190

[3] M. RIZKALLAH.*Graph based transforms for compression of new imaging modalities*, Université Rennes 1, April 2019, https://tel.archives-ouvertes.fr/tel-02285386

Articles in International Peer-Reviewed Journal

[4] P. ALLAIN, L. GUILLO, C. GUILLEMOT.*4D Anisotropic Diffusion Framework with PDEs for Light Field Regularization and Inverse Problems*, in "IEEE Transactions on Computational Imaging", May 2019, p. 1-16 [*DOI :* 10.1109/TCI.2019.2919229], https://hal.archives-ouvertes.fr/hal-02135541

[5] T. DUMAS, A. ROUMY, C. GUILLEMOT.*Context-adaptive neural network based prediction for image compression*, in "IEEE Transactions on Image Processing", August 2019, vol. 29, p. 679 - 693, https://arxiv.org/abs/1807.06244 , https://hal.archives-ouvertes.fr/hal-01841034

[6] E. DUPRAZ, A. ROUMY, T. MAUGEY, M. KIEFFER.*Rate-storage regions for Extractable Source Coding with side information*, in "Physical Communication", 2019, vol. 37, p. 1-24 [*DOI :* 10.1016/J.PHYCOM.2019.100845], https://hal.inria.fr/hal-02390517

[7] S. EVAIN, C. GUILLEMOT.*A Lightweight Neural Network for Monocular View Generation with Occlusion Handling*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", December 2019, p. 1-14, forthcoming [*DOI :* 10.1109/TPAMI.2019.2960689], https://hal.archives-ouvertes.fr/hal-02428602

[8] R. A. FARRUGIA, C. GUILLEMOT.*Light Field Super-Resolution using a Low-Rank Prior and Deep Convolutional Neural Networks*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2019, p. 1-15 [*DOI :* 10.1109/TPAMI.2019.2893666], https://hal.archives-ouvertes.fr/hal-01984843

[9] R. FARRUGIA, C. GUILLEMOT.*A Simple Framework to Leverage State-Of-The-Art Single-Image Super-Resolution Methods to Restore Light Fields*, in "Signal Processing: Image Communication", September 2019, vol. 80, p. 1-11 [*DOI :* 10.1016/J.IMAGE.2019.115638], https://hal.archives-ouvertes.fr/hal-02378049

[10] C. GALDI, V. CHIESA, C. BUSCH, P. L. CORREIA, J.-L. DUGELAY, C. GUILLEMOT.*Light Fields for Face Analysis*, in "Sensors", June 2019, vol. 19, n^{o} 12, p. 1-27 [*DOI :* 10.3390/S19122687], https://hal.archives-ouvertes.fr/hal-02157348

[11] D. GOMMELET, J. LE TANOU, A. ROUMY, M. ROPERT, C. GUILLEMOT.*Optical-Flow Based Nonlinear Weighted Prediction for SDR and Backward Compatible HDR Video Coding*, in "IEEE Transactions on Image Processing", 2019, p. 1-16, forthcoming [*DOI :* 10.1109/TIP.2019.2945685], https://hal.archives-ouvertes.fr/hal-02378053

[12] M. LE PENDU, C. GUILLEMOT, A. SMOLIC.*A Fourier Disparity Layer representation for Light Fields*, in "IEEE Transactions on Image Processing", May 2019, p. 5740 - 5753, forthcoming [*DOI :* 10.1109/TIP.2019.2922099], https://hal.archives-ouvertes.fr/hal-02130555

[13] M. RIZKALLAH, T. MAUGEY, C. GUILLEMOT.*Prediction and Sampling with Local Graph Transforms for Quasi-Lossless Light Field Compression*, in "IEEE Transactions on Image Processing", 2019, p. 1-13, forthcoming, https://hal.inria.fr/hal-02373664

[14] M. RIZKALLAH, X. SU, T. MAUGEY, C. GUILLEMOT.*Geometry-Aware Graph Transforms for Light Field Compact Representation*, in "IEEE Transactions on Image Processing", August 2019, p. 1-15, forthcoming [*DOI :* 10.1109/TIP.2019.2928873], https://hal.archives-ouvertes.fr/hal-02199839

[15] J. SHI, X. JIANG, C. GUILLEMOT.*A framework for learning depth from a flexible subset of dense and sparse light field views*, in "IEEE Transactions on Image Processing", June 2019, p. 5867-5880 [*DOI :* 10.1109/TIP.2019.2923323], https://hal.archives-ouvertes.fr/hal-02155040

International Conferences with Proceedings

[16] P. ALLAIN, L. GUILLO, C. GUILLEMOT.*Light field denoising using 4D anisotropic diffusion*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech, and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, p. 1692-1696 [*DOI :* 10.1109/ICASSP.2019.8682709], https://hal.archives-ouvertes.fr/hal-02116372

[17] J. BÉGAINT, F. GALPIN, P. GUILLOTEL, C. GUILLEMOT.*Deep frame interpolation for video compression*, in "DCC 2019 - Data Compression Conference", Snowbird, United States, IEEE, March 2019, p. 1-10 [*DOI :* 10.1109/DCC.2019.00068], https://hal.archives-ouvertes.fr/hal-02202172

[18] P. DAVID, M. LE PENDU, C. GUILLEMOT.*Sparse to Dense Scene Flow Estimation from Light Fields*, in "ICIP 2019 - IEEE International Conference on Image Processing", Taïpei, Taiwan, IEEE, September 2019, p. 1-5 [*DOI :* 10.1109/ICIP.2019.8803520], https://hal.archives-ouvertes.fr/hal-02123544

[19] E. DIB, M. LE PENDU, C. GUILLEMOT.*Light Field Compression using Fourier Disparity Layers*, in "ICIP 2019 - IEEE International Conference on Image Processing", Taipei, Taiwan, IEEE, September 2019, p. 1-5 [*DOI :* 10.1109/ICIP.2019.8803756], https://hal.archives-ouvertes.fr/hal-02130487

[20] E. DIB, M. LE PENDU, X. JIANG, C. GUILLEMOT.*Super-rays based low rank approximation for light fields compression*, in "DCC 2019 - Data Compression Conference", Snowbird, United States, IEEE, March 2019, p. 369-378 [*DOI :* 10.1109/DCC.2019.00045], https://hal.archives-ouvertes.fr/hal-02116365

[21] C. GALEA, C. GUILLEMOT.*Denoising of 3D point clouds constructed from light fields*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech, and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, p. 1882-1886 [*DOI :* 10.1109/ICASSP.2019.8683548], https://hal.archives-ouvertes.fr/hal-02116377

[22] P. GARUS, J. JOEL, T. MAUGEY, C. GUILLEMOT.*Bypassing Depth Maps Transmission For Immersive Video Coding*, in "PCS 2019 - Picture Coding Symposium", Ningbo, China, IEEE, November 2019, p. 1-5, https://hal.inria.fr/hal-02397800

[23] M. HOG, N. SABATER, C. GUILLEMOT.*Long Short Term Memory Networks For Light Field View Synthesis*, in "ICIP 2019 - IEEE International Conference on Image Processing", Taipei, Taiwan, IEEE, September 2019, p. 1-5 [*DOI :* 10.1109/ICIP.2019.8803790], https://hal.archives-ouvertes.fr/hal-02202120

[24] X. JIANG, J. SHI, C. GUILLEMOT.*A learning based depth estimation framework for 4D densely and sparsely sampled light fields*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech, and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, p. 2257-2261 [*DOI :* 10.1109/ICASSP.2019.8683773], https://hal.archives-ouvertes.fr/hal-02116375

[25] N. MAHMOUDIAN BIDGOLI, T. MAUGEY, A. ROUMY.*Evaluation framework for 360-degree visual content compression with user view-dependent transmission*, in "ICIP 2019 - IEEE International Conference on Image Processing", Taipei, Taiwan, IEEE, September 2019, p. 1-5 [*DOI :* 10.1109/ICIP.2019.8803606], https://hal.archives-ouvertes.fr/hal-02202794

[26] N. MAHMOUDIAN BIDGOLI, T. MAUGEY, A. ROUMY.*Intra-coding of 360-degree images on the sphere*, in "PCS 2019 - Picture Coding Symposium", Ningbo, France, November 2019, p. 1-5, https://hal.archives-ouvertes.fr/hal-02381193

[27] N. MAHMOUDIAN BIDGOLI, T. MAUGEY, A. ROUMY, F. NASIRI, F. PAYAN.*A geometry-aware compression of 3D mesh texture with random access*, in "PCS 2019 - Picture Coding Symposium", Ningbo, China, November 2019, p. 1-5, https://hal.archives-ouvertes.fr/hal-02380798

[28] T. MAUGEY, L. GUILLO, C. LE CAM.*FTV360: a Multiview 360°Video Dataset with Calibration Parameters*, in "MMSys 2019 - 10th ACM Multimedia Systems Conference", Amherst, United States, ACM Press, June 2019, p. 291-295 [*DOI :* 10.1145/3304109.3325815], https://hal.inria.fr/hal-02398005

[29] F. NASIRI, N. MAHMOUDIAN BIDGOLI, F. PAYAN, T. MAUGEY.*A Geometry-aware framework for compressing 3D mesh textures*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech, and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, p. 4015-4019 [*DOI :* 10.1109/ICASSP.2019.8683258], https://hal.archives-ouvertes.fr/hal-02010328

[30] C. PETIT, A. ROUMY, G. COLUCCIA, E. MAGLI.*Density evolution of Orthogonal Matching Pursuit*, in "SPARS 2019 - Signal Processing with Adaptive Sparse Structured Representations - Workshop", Toulouse, France, 2019, 1, https://hal.inria.fr/hal-02390560

[31] M. RIZKALLAH, T. MAUGEY, C. GUILLEMOT.*Graph-based Spatio-angular Prediction for Quasi-Lossless Compression of Light Fields*, in "DCC 2019 - Data Compression Conference", Snowbird, United States, IEEE, March 2019, p. 379-388 [*DOI :* 10.1109/DCC.2019.00046], https://hal.archives-ouvertes.fr/hal-02116369

National Conferences with Proceeding

[32] N. M. BIDGOLI, T. MAUGEY, A. ROUMY.*Compression de contenus 360 et transmission adaptée à la navigation de l'utilisateur*, in "GRETSI 2019 - XXVIIème Colloque francophone de traitement du signal et des images", Lille, France, August 2019, https://hal.inria.fr/hal-02390575

[33] F. NASIRI, N. M. BIDGOLI, F. PAYAN, T. MAUGEY.*Codage d'atlas de textures avec prédiction guidée par la topologie des maillages*, in "GRETSI 2019 - XXVIIème Colloque francophone de traitement du signal et des images", Lille, France, August 2019, p. 1-4, https://hal.archives-ouvertes.fr/hal-02159557

Scientific Popularization

[34] C. GUILLEMOT, A. ROUMY.*De Fourier à la compression d'images et de vidéos*, in "Interstices", October 2019, https://hal.inria.fr/hal-02336180

Project-Team STACK

Software Stack for Massively Geo-Distributed Infrastructures

IN COLLABORATION WITH: Laboratoire des Sciences du numerique de Nantes

IN PARTNERSHIP WITH:
IMT Atlantique Bretagne-Pays de la Loire

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Distributed Systems and middleware

Table of contents

Project-Team STACK

Creation of the Team: 2017 November 01, updated into Project-Team: 2019 January 01

Keywords:

Computer Science and Digital Science:

A1.1.8. - Security of architectures
A1.1.10. - Reconfigurable architectures
A1.1.13. - Virtualization
A1.3.4. - Peer to peer
A1.3.5. - Cloud
A1.3.6. - Fog, Edge
A1.5.1. - Systems of systems
A1.6. - Green Computing
A2.1.7. - Distributed programming
A2.1.10. - Domain-specific languages
A2.5.2. - Component-based Design
A2.6. - Infrastructure software
A2.6.1. - Operating systems
A2.6.2. - Middleware
A2.6.3. - Virtual machines
A2.6.4. - Ressource management
A3.1.2. - Data management, quering and storage
A3.1.3. - Distributed data
A3.1.8. - Big data (production, storage, transfer)
A4.1. - Threat analysis
A4.4. - Security of equipment and software
A4.9. - Security supervision

Other Research Topics and Application Domains:

B2. - Health
B4. - Energy
B4.5.1. - Green computing
B5.1. - Factory of the future
B6.3. - Network functions
B6.4. - Internet of things
B6.5. - Information systems
B7. - Transport and logistics
B8. - Smart Cities and Territories

1. Team, Visitors, External Collaborators

Research Scientist

Shadi Ibrahim [Inria, Researcher]

Faculty Members

Adrien Lebre [Team leader, IMT Atlantique, Professor, HDR]
Hélène Coullon [IMT Atlantique, Inria Chair, Associate Professor]
Thomas Ledoux [IMT Atlantique, Associate Professor, HDR]
Jean-Marc Menaud [IMT Atlantique, Professor, HDR]
Jacques Noyé [IMT Atlantique, Associate Professor]
Mario Südholt [IMT Atlantique, Professor, HDR]

Technical Staff

Ronan-Alexandre Cherrueau [Inria, Engineer]
Marie Delavergne [Inria, Engineer, until May 2019]
Brice Nédelec [Sigma, Engineer, from Sep 2019]
Jonathan Pastor [IMT Atlantique, Engineer]
Dimitri Pertin [IMT Atlantique, Engineer, from Sep 2019 until Oct 2019]
Rémy Pottier [IMT Atlantique, Engineer, from Nov 2019]
Javier Rojas Balderrama [Inria, Engineer]
Charlène Servantie [IMT Atlantique, Engineer, from Jul 2019]
Matthieu Simonin [Inria, Engineer]

PhD Students

Mohamed Abderrahim [Inria, PhD Student, until Feb 2019]
Mohammad-Mahdi Bazm [Orange Labs, PhD Student, until Jun 2019]
Maxime Belair [Orange Labs, PhD Student, granted by CIFRE]
Fatima Zahra Boujdad [Univ de Nantes, PhD Student, until Nov 2019]
Emile Cadorel [Armines, PhD Student]
Maverick Chardet [Inria, PhD Student]
Jad Darrous [Inria, PhD Student, from Oct 2019]
Marie Delavergne [Inria, PhD Student, from Oct 2019]
David Espinel [Orange Labs, PhD Student, granted by CIFRE]
Wilmer Edicson Garzon Alfonso [IMT Atlantique, PhD Student, from Jun 2019]
Thuy Linh Nguyen [Inria, PhD Student, until Jun 2019]
Dimitri Saingre [Armines, PhD Student]
Sirine Sayadi [IMT Atlantique, PhD Student]
Yewan Wang [Orange Labs, PhD Student, until Oct 2019, granted by CIFRE]

Post-Doctoral Fellows

David Guyon [Inria, Post-Doctoral Fellow, from Apr 2019]
Thomas Lambert [Inria, Post-Doctoral Fellow]
Hamza Sahli [Inria, Post-Doctoral Fellow]
Simon Robillard [IMT Atlantique, Post-Doctoral Fellow, from Jul 2019]
Alexandre Van Kempen [Inria, Post-Doctoral Fellow, until Jun 2019]

Visiting Scientist

Twinkle Jain [Northeastern University, from May 2019 until Jul 2019]

Administrative Assistant

Anne-Claire Binétruy [Inria, Administrative Assistant]

2. Overall Objectives

2.1. STACK in a Nutshell

The STACK team addresses challenges related to the management and advanced usages of Utility Computing infrastructures (i.e., Cloud, Fog, Edge, and beyond). More specifically, the team is interested in delivering appropriate system abstractions to operate and use massively geo-distributed ICT infrastructures, from the lowest (system) levels to the highest (application development) ones, and addressing crosscutting dimensions such as energy or security. These infrastructures are critical for the emergence of new kinds of applications related to the digitalization of the industry and the public sector (a.k.a. the Industrial and Tactile Internet).

2.2. Toward a STACK for Geo-Distributed Infrastructures

With the advent of Cloud Computing, modern applications have been developed on top of advanced software stacks composed of low-level system mechanisms, advanced middleware and software abstractions. While each of these layers has been designed to enable developers to efficiently use ICT resources without dealing with the burden of the underlying infrastructure aspects, the complexity of the resulting software stack has become a key challenge. As an example, Map/Reduce frameworks such as Hadoop have been developed to benefit from cpu/storage capacities of distinct servers. Running such frameworks on top of a virtualized cluster (*i.e.*, in a Cloud) can lead to critical situations if the resource management system decides to consolidate all the VMs on the same physical machine [101]. In other words, self-management decisions taken in isolation at one level (infrastructure, middleware, or application) may indirectly interfere with the decision taken by another layer, and globally affect the performance of the whole stack. Considering that geo-distributed ICT infrastructures significantly differ from the Cloud Computing ones regarding heterogeneity, resiliency, and the potential massive distribution of resources and networking environments [63], [96], we can expect that the complexity of the software stacks is going to increase. Such an assumption can be illustrated, for instance, by the sotfware architecture proposed in 2016 by the ETSI Mobile edge computing Industry Specification Group [88]. This architecture is structured around a new layer in charge of orchestrating distinct independent cloud systems, *a.k.a.* Virtual Infrastructure Managers (VIMs) in their terminology. By reusing VIMs, ETSI targets an edge computing resource management that behaves in the same fashion as Cloud Computing ones. While mitigating development requirements, such a proposal hides all management decisions that might be taken in the VIM of one particular site and thus may lead to conflicting decisions and consequently to non-desired states overall.

Through the STACK team, we propose to investigate the sotfware stack challenge as a whole. We claim it is the only way to limit as much as possible the complexity of the next generation software stack of geo-distributed ICT infrastructures. To reach our goal, we will identify major building blocks that should compose such a software stack, how they should be designed (*i.e.*, from the internal algorithms to the APIs they should expose), and finally how they should interact with each other.

Delivering such a software stack is an ambitious objective that goes beyond the activities of one research group. However, our expertise, our involvements in different consortiums (such as OpenStack) as well as our participation to different collaborative projects enable STACK members to contribute to this challenge in terms of architecture models, distributed system mechanisms and software artefacts, and finally, guideline reports on opportunities and constraints of geo-distributed ICT infrastructures.

3. Research Program

3.1. Overview

STACK research activities have been organized around four research topics. The two first ones are related to the resource management mechanisms and the programming support that are mandatory to operate and use ICT geo-distributed resources (compute, storage, network). They are transverse to the System/Middleware/Application layers, which generally composed a software stack, and nurture each other (*i.e.*, the resource management mechanisms will leverage abstractions/concepts proposed by the programming support axis and reciprocally). The third and fourth research topics are related to the Energy and Security dimensions (both also crosscutting the three software layers). Although they could have been merged with the two first axes, we identified them as independent research directions due to their critical aspects with respect to the societal challenges they represent. In the following, we detail the actions we plan to do in each research direction.

3.2. Resource Management

The challenge in this axis is to identify, design or revise mechanisms that are mandatory to operate and use a set of massively geo-distributed resources in an efficient manner [50]. This covers considering challenges at the scale of nodes, within one site (*i.e.*, one geographical location) and throughout the whole geo-distributed ICT infrastructure. It is noteworthy that the network community has been investigating similar challenges for the last few years [69]. To benefit from their expertise, in particular on how to deal with intermittent networks, STACK members have recently initiated exchanges and collaborative actions with some network research groups and telcos (see Sections 8.1 and 9.1). We emphasize, however, that we do not deliver contributions related to network equipments/protocols. The scientific and technical achievements we aim to deliver are related to the (distributed) system aspects.

3.2.1. Performance Characterization of Low-Level Building Blocks

Although Cloud Computing has enabled the consolidation of services and applications into a subset of servers, current operating system mechanisms do not provide appropriate abstractions to prevent (or at least control) the performance degradation that occurs when several workloads compete for the same resources [101]. Keeping in mind that server density is going to increase with physical machines composed of more and more cores and that applications will be more and more data intensive, it is mandatory to identify interferences that appear at a low level on each dimension (compute, memory, network, and storage) and propose countermeasures. In particular, previous studies [101], [61] on pros and cons of current technologies – virtual machines (VMs) [75], [83], containers and microservices – which are used to consolidate applications on the same server, should be extended: In addition to evaluating the performance we can expect from each of these technologies on a single node, it is important to investigate interferences that may result from cross-layer and remote communications [102]. We will consider in particular all interactions related to geo-distributed systems mechanisms/services that are mandatory to operate and use geo-distributed ICT infrastructures.

3.2.2. Geo-Distributed System Mechanisms

Although several studies have been highlighting the advantages of geo-distributed ICT infrastructures in various domains (see Section 3.), progress on how to operate and use such infrastructures is marginal. Current solutions [32] [33] are rather close to the initial Cisco Fog Computing proposal that only allows running domain-specific applications on edge resources and centralized Cloud platforms [40] (in other words, these solutions do not allow running stateful workloads in isolated environments such as containers or VMs). More recently, solutions leveraging the idea of federating VIMs (as the aforementioned ETSI MEC proposal [88]) have been proposed. ONAP [95], an industry-driven solution, enables the orchestration and automation of virtual network functions across distinct VIMs. From the academic side, FogBow [42] aims to support federations of Infrastructure-as-a-Service (IaaS) providers. Finally, NIST initiated a collaborative effort with IEEE to advance Federated Cloud platforms through the development of a conceptual architecture and a vocabulary[0]. Although all these projects provide valuable contributions, they face the aforementioned orchestration limitations (*i.e.*, they do not manage decisions taken in each VIM). Moreover, they all have been designed by only considering the developer/user's perspective. They provide abstractions to manage the life cycle of geo-distributed applications, but do not address administrative requirements.

[0] https://collaborate.nist.gov/twiki-cloud-computing/bin/view/CloudComputing/FederatedCloudPWGFC (Dec 2018).

To cope with specifics of Wide-Area networks while delivering most features that made Cloud Computing solutions successful also at the edge, our community should first identify limitations/drawbacks of current resource management system mechanisms with respect to the Fog/Edge requirements and propose revisions when needed [68], [81].

To achieve this aim, STACK members propose to conduct first a series of studies aiming at understanding the software architecture and footprint of major services that are mandatory for operating and using Fog/Edge infrastructures (storage backends, monitoring services, deployment/reconfiguration mechanisms, etc.). Leveraging these studies, we will investigate how these services should be deployed in order to deal with resources constraints, performance variability, and network split brains. We will rely on contributions that have been accomplished in distributed algorithms and self-* approach for the last decade. In the short and medium term, we plan to evaluate the relevance of NewSQL systems [53] to store internal states of distributed system mechanisms in an edge context, and extend our proposals on new storage backends such as key/value stores [52], [94], and burst buffers [103]. We also plan to conduct new investigations on data-stream frameworks for Fog and Edge infrastructures [47]. These initial contributions should enable us to identify general rules to deliver other advanced system mechanisms that will be mandatory at the higher levels in particular for the deployment and reconfiguration manager in charge of orchestrating all resources.

3.2.3. Capacity Planning and Placement Strategies

An objective shared by users and providers of ICT infrastructures is to limit as much as possible the operational costs while providing the expected and requested quality of service (QoS). To optimize this cost while meeting QoS requirements, data and applications have to be placed in the best possible way onto physical resources according to data sources, data types (stream, graphs), application constraints (real-time requirements) and objective functions. Furthermore, the placement of applications must evolve through time to cope with the fluctuations in terms of application resource needs as well as the physical events that occur at the infrastructure level (resource creation/removals, hardware failures, etc.). This placement problem, *a.k.a.* the deployment and reconfiguration challenge as it will be described in Section 3.3, can be modelized in many different ways, most of the time by multi-dimensional and multi-objective bin-packing problems or by scheduling problems which are known to be difficult to solve. Many studies have been performed, for example, to optimize the placement of virtual machines onto ICT infrastructures [77]. STACK will inherit the knowledge acquired through previous activities in this domain, particularly its use of constraint programming strategies in autonomic managers [73], [72], relying on MAPE (monitor, analyze, plan, and execute) control loops. While constraint programming approaches are known to hardly scale, they enable the composition of various constraints without requiring to change heuristic algorithms each time a new constraint has to be considered [71]. We believe it is a strong advantage to deal with the diversity brought by geo-distributed ICT infrastructures. Moreover, we have shown in previous work that decentralized approaches can tackle the scalability issue while delivering placement decisions good enough and sometimes close to the optimal [87].

Leveraging this expertise, we propose, first, to identify new constraints raised by massively geo-distributed infrastructures (*e.g.*, data locality, energy, security, reliability and the heterogeneity and mobility of the underlying infrastructure). Based on this preliminary study, we will explore new placement strategies not only for computation sandboxes but for data (location, replication, streams, etc.) in order to benefit from the geo-distribution of resources and meet the required QoS.These investigations should lead to collaborations with operational research and optimization groups such as TASC, another research group from IMT Atlantique.

Second, we will leverage contributions made on the previous axis “Performance Characterization of Low-Level Building Blocks” to determine how the deployment of the different units (software components and data sets) should be executed in order to reduce as much as possible the time to reconfigure the system (*i.e.*, the *Execution* phase in the control loop). In some recent work [83], we have shown that the provisioning of a new virtual machine should be done carefully to mitigate boot penalties. More generally, proposing an efficient action plan for the *Execution* phase will be a major point as Wide-Area-Network specifics may lead to significant delays, in particular when the amount of data to be manipulated is important.

Finally, we will investigate new approaches to decentralize the placement process while considering the geo-distributed context. Among the different challenges to address, we will study how a combination of autonomic managers, at both the infrastructure and application levels [60], could be proposed in a decentralized manner. Our first idea is to geo-distribute a fleet of small control loops over the whole infrastructure. By improving the locality of data collection and weakening combinatorics, these loops would allow the system to address responsiveness and quality expectations.

3.3. Programming Support

We pursue two main research directions relative to new programming support: first, developing new programming models with appropriate support in existing languages (libraries, embedded DSLs, etc.) and, second, providing new means for deployment and reconfiguration in geo-distributed ICT environments, principally supporting the mapping of software onto the infrastructure. For both directions two levels of challenges are considered. On the one hand, the *generic* level refers to efforts on programming support that can be applied to any kind of distributed software, application or system. On this level, contributions could thus be applied to any of the three layers addressed by STACK (*i.e.*, system, middleware or application). On the other hand, the corresponding generic programming means may not be appropriate in practice (*e.g.*, requirements for more dedicated support, performance constraints, etc.), even if they may lead to interesting general properties. For this reason, a *specific* level is also considered. This level could be based on the generic one but addresses specific cases or domains.

3.3.1. Programming Models and Languages Extensions

The current landscape of programming support for cloud applications is fragmented. This fragmentation is based on apparently different needs for various kinds of applications, in particular, web-based, computation-based, focusing on the organization of the computation, and data-based applications, within the last case a quite strong dichotomy between applications considering data as sets or relations, close to traditional database applications and applications considering data as real-time streams. This has led to various programming models, in a loose sense, including for instance microservices, graph processing, dataflows, streams, etc. These programming models have mostly been offered to the application programmer in the guise of frameworks, each offering subtle variants of the programming models with various implementation decisions favoring particular application and infrastructure settings. Whereas most frameworks are dedicated to a given programming model, *e.g.*, basic Pregel [82], Hive [97], Hadoop [98], some of them are more general-purpose through the provision of several programming models, *e.g.*, Flink [46] and Spark [79]. Finally, some dedicated language support has been considered for some models (*e.g.*, the language SPL underlying IBM Streams [74]) as well as core languages and calculi (*e.g.*, [43], [92]).

This situation raises a number of challenges on its own, related to a better structuring of the landscape. It is necessary to better understand the various programming models and their possible relations, with the aim of facilitating, if not their complete integration, at least their composition, at the conceptual level but also with respect to their implementations, as specific languages and frameworks.

Switching to massively geo-distributed infrastructures adds to these challenges by leading to a new range of applications (*e.g.*, smart-* applications) that, by nature, require mixing these various programming models, together with a much more dynamic management of their runtime.

In this context, STACK would like to explore two directions:

- First, we propose to contribute to generic programming models and languages to address composability of different programming models [55]. For example, providing a generic stream data processing model that can operate under both data stream [46] and operation stream [104] modes, thus streams can be processed in micro batches to favour high throughput or record by record to sustain low latency. Software engineering properties such as separation of concerns and composition should help address such challenges [35], [93]. They should also facilitate the software deployment and reconfiguration challenges discussed below.

- Second, we plan to revise relevant programming models, the associated specific languages, and their implementation according to the massive geo-distribution of the underlying infrastructure, the data sources, and application end-users. For example, although SPL is extensible and distributed, it has been designed to run on multi-cores and clusters [74]. It does not provide the level of dynamicity required by geo-distributed applications (*e.g.*, to handle topology changes, loss of connectivity at the edge, etc.). Moreover, as more network data transfers will happen within a massively geo-distributed infrastructure, correctness of data transfers should be guaranteed. This has potential impact from the programming models to their implementations.

3.3.2. Deployment and Reconfiguration Challenges

The second research direction deals with the complexity of deploying distributed software (whatever the layer, application, middleware or system) onto an underlying infrastructure. As both the deployed pieces of software and the infrastructures addressed by STACK are large, massively distributed, heterogeneous and highly dynamic, the deployment process cannot be handled manually by developers or administrators. Furthermore, and as already mentioned in Section 3.2, the initial deployment of some distributed software will evolve through time because of the dynamicity of both the deployed software and the underlying infrastructures. When considering reconfiguration, which encompasses deployment as a specific case, the problem becomes more difficult for two main reasons: (1) the current state of both the deployed software and the infrastructure has to be taken into account when deciding on a reconfiguration plan, (2) as the software is already running the reconfiguration should minimize disruption time, while avoiding inconsistencies [80], [85]. Many deployment tools have been proposed both in academia and industry [57]. For example, Ansible [0], Chef [0] and Puppet [0] are very well-known generic tools to automate the deployment process through a set of batch instructions organized in groups (*e.g.*, *playbooks* in Ansible). Some tools are specific to a given environment, like Kolla to deploy OpenStack, or the embedded deployment manager within Spark. Few reconfiguration capabilities are available in production tools such as *scaling* and *restart* after a fault [0] [0]. Academia has contributed to generic deployment and reconfiguration models. Most of these contributions are component-based. Component models divide a distributed software as a set of component instances (or modules) and their assembly, where components are connected through well defined interfaces [93]. Thus, modeling the reconfiguration process consists in describing the life cycle of different components and their interactions. Most component-based approaches offer a fixed life cycle, *i.e.*, identical for any component [62]. Two main contributions are able to customize life cycles, Fractal [45], [38] and its evolutions [35], [36], [59], and Aeolus [54]. In Fractal, the *control* part of a component (*e.g.*, its life cycle) is modeled itself as a component assembly that is highly flexible. Aeolus, on the other hand, offers a finer control on both the evolution and the synchronization of the deployment process by modeling each component life cycle with a finite state machine.

A reconfiguration raises at least five questions, all of them are correlated: (1) *why software has to be reconfigured?* (monitoring, modeling and analysis) (2) *what should be reconfigured?* (software modeling and analysis), (3) *how should it be reconfigured?* (software modeling and planning decisions), (4) *where should it be reconfigured?* (infrastructure modeling and planning decisions), and (5) *when to reconfigure it?* (scheduling algorithms). STACK will contribute to all aspects of a reconfiguration process as described above. However, according to the expertise of STACK members, we will focus mainly on the three first questions: *why*, *what* and *how*, leaving questions *where* and *when* to collaborations with operational research and optimization teams.

First of all, we would like to investigate *why software has to be reconfigured?* Many reasons could be mentioned, such as hardware or software fault tolerance, mobile users, dynamicity of software services, etc. All those reasons are related somehow to the Quality of Service (QoS) or the Service Level Agreement (SLA) between the user and the Cloud provider. We first would like to explore the specificities of QoS and SLAs

[0] https://www.ansible.com/
[0] https://www.chef.io/chef/
[0] https://puppet.com/
[0] https://kubernetes.io/
[0] https://jujucharms.com/

in the case of massively geo-distributed ICT environments [89]. By being able to formalize this question, analyzing the requirement of a reconfiguration will be facilitated.

Second, we think that four important properties should be enhanced when deploying and reconfiguring models in massively geo-distributed ICT environments. First, as low-latency applications and systems will be subject to deployment and reconfiguration, the performance and the ability to scale are important. Second, as many different kinds of deployments and reconfigurations will concurrently hold within the infrastructure, processes have to be reliable, which is facilitated by a fine-grained control of the process. Finally, as many different software elements will be subject to deployment and reconfiguration, common generic models and engines for deployment and reconfiguration should be designed [44]. For these reasons, we intend to go beyond Aeolus by: first, leveraging the expression of parallelism within the deployment process, which should lead to better performance; second, improving the separation of concerns between the component developer and the reconfiguration developer; third, enhancing the possibility to perform concurrent and decentralized reconfigurations.

Research challenges relative to programming support have been presented above. Many of these challenges are related, in different manners, to the resource management level of STACK or to crosscutting challenges, *i.e.*, energy and security. First, one can notice that any programming model or deployment and reconfiguration implementation should be based on mechanisms related to resource management challenges. For this reason, all challenges addressed within this section are linked with lower level building blocks presented in Section 3.2. Second, as detailed above, deployment and reconfiguration address at least five questions. The question *what?* is naturally related to programming support. However, questions *why*, *how?*, *where?* and *when?* are also related to Section 3.2, for example, to monitoring and capacity planning. Moreover, regarding the deployment and reconfiguration challenges, one can note that the same goals recursively happen when deploying the control building blocks themselves (bootstrap issue). This comforts the need to design generic deployment and reconfiguration models and frameworks. These low-level models should then be used as back-ends to higher-level solutions. Finally, as *energy* and *security* are crosscutting themes within the STACK project, many additional energy and security considerations could be added to the above challenges. For example, our deployment and reconfiguration frameworks and solutions could be used to guarantee the deployment of end-to-end security policies or to answer specific energy constraints [70] as detailed in the next section.

3.4. Energy

The overall electrical consumption of DCs grows according to the demand of Utility Computing. Considering that the latter has been continuously increasing since 2008, the energy footprint of Cloud services overall is nowadays critical with about 91 billion kilowatt-hours of electricity [91]. Besides the ecological impact, the energy consumption is a predominant criterion for providers since it determines a large part of the operational cost of their infrastructure. Among the different appraoches that have been investigated to reduce the energy footprint, some studies have been ingestigating the use of renewable energy sources to power microDCs [64]. Workload distribution for geo-distributed DCs is also another promising approach [66], [78], [99]. Our research will extend these results with the ultimate goal of considering the different opportunities to control the energy footprint across the whole stack (hardware and software opportunities, renewable energy, thermal management, etc.). In particular, we identified several challenges that we will address in this context within the STACK framework.

First, we propose to evaluate the energy efficiency of low-level building blocks, from the viewpoints of computation (VMs, containers, microkernel, microservices) [58] and data (hard drives, SSD, in-memory storage, distributed file systems). For computations, in the continuity of our previous work [56], [73], we will investigate workload placement policies according to energy (minimizing energy consumption, power capping, thermal load balancing, etc.). Regarding the data dimension, we will investigate, in particular, the trade-offs between energy consumption and data availability, durability and consistency [51], [94]. Our ambition is to propose an adaptive energy-aware data layout and replication scheme to ensure data availability with minimal energy consumption. It is noteworthy that these new activities will also consider our previous work on DCs partially powered by renewable energy (see the SeDuCe project, in Section 6.7), with the ultimate goal of reducing the CO_2 footprint.

Second, we will complete current studies to understand pros and cons of massively geo-distributed infrastructures from the energy perspective. Addressing the energy challenge is a complex task that involves considering several dimensions such as the energy consumption due to the physical resources (CPU, memory, disk, network), the performance of the applications (from the computation and data viewpoints), and the thermal dissipation caused by air conditioning in each DC. Each of these aspects can be influenced by each level of the software stack (*i.e.*, low-level building blocks, coordination and autonomous loops, and finally application life cycle). In previous projects, we have studied and modeled the consumption of the main components, notably the network, as part of a single microDC. We plan to extend these models to deal with geo-distribution. The objective is to propose models that will enable us to refine our placement algorithms as discussed in the next paragraph. These models should be able to consider the energy consumption induced by all WAN data exchanges, including site-to-site data movements as well as the end users' communications for accessing virtualized resources.

Third, we expect to implement green-energy-aware balancing strategies, leveraging the aforementioned contributions. Although the infrastructures we envision increase complexity (because WAN aspects should also be taken into account), the geo-distribution of resources brings several opportunities from the energy viewpoint. For instance, it is possible to define several workload/data placement policies according to renewable energy availability. Moreover, a tightly-coupled software stack allows users to benefit from such a widely distributed infrastructure in a transparent way while enabling administrators to balance resources in order to benefit from green energy sources when available. An important difficulty, compared to centralized infrastructures, is related to data sharing between software instances. In particular, we will study issues raised by the distribution and replication of services across several microDCs. In this new context, many challenges must be addressed: where to place the data (Cloud, Edge) in order to mitigate dat a movements? What is the impact in terms of energy consumption, network and response time of these two approaches? How to manage the consistency of replicated data/services? All these aspects must be studied and integrated into our placement algorithms.

Fourth, we will investigate the energy footprint of the current techniques that address failure and performance variability in large-scale systems. For instance, *stragglers* (*i.e.*, tasks that take a significantly longer time to finish than the normal execution time) are natural results of performance variability, they cause extra resource and energy consumption. Our goal is to understand the energy overhead of these techniques and introduce new handling techniques that take into consideration the energy efficiency of the platform [86].

Finally, in order to answer specific energy constraints, we want to reify energy aspects at the application level and propose a metric related to the use of energy (Green SLA [34]), for example to describe the maximum allowed CO_2 emissions of a Fog/Edge service. Unlike other approaches [67], [39], [65] that attempt to identify the best trade-off, we want to offer to developers/end-users the opportunity to select the best choice between application performance, correctness and energy footprint. Such a capability will require reifying the energy dimension at the level of big-data and interactive applications. Besides, with the emergence of renewable energy (*e.g.*, solar panels for microDC), investigating the energy consumption *vs* performance trade-off [70] and the smart usage of green energy for ICT geo-distributed services seems promising. For example, we want to offer the opportunity to developers/end-users to control the scaling of the applications based on this trade-off instead of current approaches that only considered application load. Providing such a capability will also require appropriate software abstractions.

3.5. Security

Because of its large size and complex software structure, geo-distributed applications and infrastructures are particularly exposed to security and privacy issues [90]. They are subject to numerous security vulnerabilities that are frequently exploited by malicious attackers in order to exfiltrate personal, institutional or corporate data. Securing these systems require security and privacy models and corresponding techniques that are applicable at all software layers in order to guard interactions at each level but also between levels. However, very few security models exist for the lower layers of the software stack and no model enables the handling of interactions involving the complete software stack. Any modification to its implementation, deployment status, configuration, etc., may introduce new or trigger existing security and privacy issues. Finally, applications that execute on top of the software stack may introduce security issues or be affected by vulnerabilities of the stack. Overall, security and privacy issues are therefore interdependent with all other activities of the STACK team and constitute an important research topic for the team.

As part of the STACK activities, we consider principally security and privacy issues related to the vertical and horizontal compositions of software components forming the software stack and the distributed applications running on top of it. Modifications to the *vertical composition* of the software stack affect different software levels at once. As an example, side-channel attacks often target virtualized services (*i.e.*, services running within VMs); attackers may exploit insecure hardware caches at the system level to exfiltrate data from computations at the higher level of VM services [84], [100]. Security and privacy issues also affect *horizontal compositions*, that is, compositions of software abstractions on one level: most frequently horizontal compositions are considered on the level of applications/services but they are also relevant on the system level or the middleware level, such as compositions involving encryption and database fragmentation services.

The STACK members aim at addressing two main research issues: enabling full-stack (vertical) security and per-layer (horizontal) security. Both of these challenges are particularly hard in the context of large geo-distributed systems because they are often executed on heterogeneous infrastructures and are part of different administrative domains and governed by heterogeneous security and privacy policies. For these reasons they typically lack centralized control, are frequently subject to high latency and are prone to failures.

Concretely, we will consider two classes of security and privacy issues in this context. First, on a general level, we strive for a method for the programming and reasoning about compositions of security and privacy mechanisms including, but not limited to, encryption, database fragmentation and watermarking techniques. Currently, no such general method exists, compositions have only been devised for specific and limited cases, for example, compositions that support the commutation of specific encryption and watermarking techniques [76], [48]. We provided preliminary results on such compositions [49] and have extended them to biomedical, notably genetic, analyses in the e-health domain [41]. Second, on the level of security and privacy properties, we will focus on isolation properties that can be guaranteed through vertical and horizontal composition techniques. We have proposed first results in this context in form of a compositional notion of distributed side channel attacks that operate on the system and middleware levels [37].

It is noteworthy that the STACK members do not have to be experts on the individual security and privacy mechanisms, such as watermarking and database fragmentation. We are, however, well-versed in their main properties so that we can integrate them into our composition model. We also interact closely with experts in these techniques and the corresponding application domains, notably e-health for instance, in the context of the PRIVGEN project[0], see Section 9.1.

More generally, we highlight that security issues in distributed systems are very closely related to the other STACK challenges, dimensions and research directions. Guaranteeing security properties across the software stack and throughout software layers in highly volatile and heterogeneous geo-distributed systems is expected to harness and contribute results to the self-management capabilities investigated as part of the team's resource management challenges. Furthermore, security and privacy properties are crosscutting concerns that are intimately related to the challenges of application life cycle management. Similarly, the security issues are also closely related to the team's work on programming support. This includes new means for programming, notably in terms of event and stream programming, but also the deployment and reconfiguration challenges, notably concerning automated deployment. As a crosscutting functionality, the security challenges introduced above must be met in an integrated fashion when designing, constructing, executing and adapting distributed applications as well as managing distributed resources.

4. Application Domains

[0]Privacy-preserving sharing and processing of genetic data, https://privgen.cominlabs.u-bretagneloire.fr/fr

4.1. Overview

Supporting industrial actors and open-source communities in building an advanced software management stack is a key element to favor the advent of new kinds of information systems as well as web applications. Augmented reality, telemedecine and e-health services, smart-city, smart-factory, smart-transportation and remote security applications are under investigations. Although, STACK does not intend to address directly the development of such applications, understanding their requirements is critical to identify how the next generation of ICT infrastructures should evolve and what are the appropriate software abstractions for operators, developers and end-users. STACK team members have been exchanging since 2015 with a number of industrial groups (notably Orange Labs and Airbus), a few medical institutes (public and private ones) and several telecommunication operators in order to identify both opportunities and challenges in each of these domains, described hereafter.

4.2. Industrial Internet

The Industrial Internet domain gathers applications related to the convergence between the physical and the virtual world. This convergence has been made possible by the development of small, lightweight and cheap sensors as well as complex industrial physical machines that can be connected to the Internet. It is expected to improve most processes of daily life and decision processes in all societal domains, affecting all corresponding actors, be they individuals and user groups, large companies, SMEs or public institutions. The corresponding applications cover: the improvement of business processes of companies and the management of institutions (*e.g.*, accounting, marketing, cloud manufacturingi, etc.); the development of large “smart” applications handling large amounts of geo-distributed data and a large set of resources (video analytics, augmented reality, etc.); the advent of future medical prevention and treatment techniques thanks to the intensive use of ICT systems, etc. We expect our contributions will favor the rise of efficient, correct and sustainable massively geo-distributed infrastructures that are mandatory to design and develop such applications.

4.3. Internet of Skills.

The Internet of Skills is an extension of the Industrial Internet to human activities. It can be seen as the ability to deliver physical experiences remotely (*i.e.*, via the Tactile Internet). Its main supporters advocate that it will revolutionize the way we teach, learn, and interact with pervasive resources. As most applications of the Internet of Skills are related to real time experiences, latency may be even more critical than for the Industrial Internet and raise the locality of computations and resources as a priority. In addition to identifying how Utility Computing infrastructures can cope with this requirement, it is important to determine how the quality of service of such applications should be defined and how latency and bandwidth constraints can be guaranteed at the infrastructure level.

4.4. e-Health

The e-Health domain constitutes an important societal application domain of the two previous areas. The STACK teams is investigating distribution, security and privacy issues in the fields of systems and personalized (aka. precision) medicine. The overall goal in these fields is the development of medication and treatment methods that are tailored towards small groups or even individual patients.

We are working, as part of the ongoing PrivGen CominLabs collaborative project on new means for the sharing of genetic data and applications in the Cloud. More generally, we are applying and developing corresponding techniques for the medical domains of genomics, immunobiology and transplantalogy in the international network SHLARC and the regional networks SysMics and Oncoshare: there, we investigate how to secure and preserve privacy if potentially sensitive personal data is moved and processed by distributed biomedical analyses.

We are also involved in the SyMeTRIC regional initiative where preliminary studies have been conducted in order to build a common System Medicine computing infrastructure to accelerate the discovery and validation of bio-markers in the fields of oncology, transplantation, and chronic cardiovascular diseases. The challenges were related to the need of being able to perform analyses on data that cannot be moved between distinct locations.

The STACK team will continue to contribute to the e-Health domain by harnessing advanced architectures, applications and infrastructures for the Fog/Edge.

4.5. Network Virtualization and Mobile Edge Services.

Telecom operators have been among the first to advocate the deployment of massively geo-distributed infrastructures, in particular through working groups such as Mobile Edge Computing at the European Telecommunication Standards Institute[0]. The initial reason is that geo-distributed infrastructures will enable Telecom operators to virtualize a large part of their resources and thus reduce capital and operational costs. As an example, we are investigating through the I/O Lab, the joint lab between Orange and Inria, how can a Centralized Radio Access Networks (*a.k.a.* C-RAN or Cloud-RAN) be supported for 5G networks. We highlight that our expertise is not on the network side but rather on where and how we can deploy, allocate and reconfigure software components, which are mandatory to operate a C-RAN infrastructure, in order to guarantee the quality of service expected by the end-users. Finally, working with actors from the network community is a valuable advantage for a distributed system research group such as STACK. Indeed, achievements made within one of the two communities serve the other.

5. Highlights of the Year

5.1. Highlights of the Year

Regarding scientific results, the team has produced a number of outstanding results on the management of resources and data in large-scale infrastructures, notably on how speeding up VM and Docker boot time by reducing the I/O operations [16], on how to place container images across edge servers in such a way that an image can be retrieved from any edge server fast and in a predictable time [14], and how the placement challenge of data and computations accross multiple sites can be addressed by using Constraint Programming techniques in a general manner [22].

We also deliver two other important contributions. In the first one, we propose an efficient graph partitioning method named Geo-Cut, which takes both the cost and performance objectives into consideration for large graph processing in geo-distributed DCs [8]. In the second one, we propose a model and a first implementation of a simulator to compare the energy footprint of different cloud architectures (single sites vs fully decentrlaized) [3].

On the software side, the team has pursued its efforts on the development of the EnosLib library and the resulting artifacts to help researchers perform experiment campaigns: https://discovery.gitlabpages.inria.fr/enoslib/theyuseit.html. We would like also to point it out the development of the field of dynamic reconfiguration of distributed software systems, in particular through the Concerto and Mad softwares: http://helene-coullon.fr/verdi/page/software/

On the platform side, the deployment of the SeDuCe testbed that allows researchers to investigate energy concerns in data-centers thanks to a numerous of energy sensors deployed across the dedicated facility is now fully operational: https://seduce.fr. Moreover, the team is still strongly involved in the different actions that aim to setup the SILECS platform.

[0] http://www.etsi.org/news-events/news/1078-2016-04-etsi-mobile-edge-computing-publishes-foundation-specifications.

5.1.1. Awards

In 2019, the team has received two individual award:

- **Outstanding Leadership Award** Shadi Ibrahim received an outstanding leadership award as program chair of the SmartData-2019 (http://cse.stfx.ca/~cybermatics/2019/smartdata/).
- **Best Tech Pitch** Hélène Coullon received the best tech pitch award at the IMT 5G event from a jury composed of both academic experts in 5G and experts from the Qualcomm company. Moreover, a grant has been awarded by France Brevet to Hélène Coullon to push further her efforts on Fog and Edge computing.

We would like also to highlight two other elements that underline the visibility and recognition of the team nationally and internationally. First, Thomas Ledoux became head of the teaching chair "ArchOps : architecture, déploiement et administration des infrastructures IT agiles" supported by Bodet Software. The ArchOps chair aims to develop skills in the design of distributed software architectures for engineering students at IMT Atlantique. Second, Shadi Ibrahim and Hélène Coullon, two members of the team, act as program track chairs of 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGrid 2020), a major conference in the area of distributed systems.

6. New Software and Platforms

6.1. MAD

Madeus Application Deployer

KEYWORDS: Automatic deployment - Distributed Software - Component models - Cloud computing

SCIENTIFIC DESCRIPTION: MAD is a Python implementation of the Madeus deployment model for multi-component distributed software. Precisely, it allows to: 1. describe the deployment process and the dependencies of distributed software components in accordance with the Madeus model, 2. describe an assembly of components, resulting in a functional distributed software, 3. automatically deploy the component assembly of distributed software following the operational semantics of Madeus.

RELEASE FUNCTIONAL DESCRIPTION: Initial submission with basic functionalities of MAD

NEWS OF THE YEAR: Operational prototype.

- Participants: Christian Pérez, Dimitri Pertin, Hélène Coullon and Maverick Chardet
- Partners: IMT Atlantique - LS2N - LIP
- Contact: Hélène Coullon
- Publications: Madeus: A formal deployment model - Behavioral interfaces for reconfiguration of component models

6.2. Nitro

KEYWORDS: Cloud storage - Virtual Machine Image - Geo-distribution

SCIENTIFIC DESCRIPTION: Nitro is a storage system that is designed to work in geo-distributed cloud environments (i.e., over WAN) to efficiently manage Virtual Machine Images (VMIs).

Nitro employs fixed-size deduplication to store VMIs. This technique contributes to minimizing the network cost. Also, Nitro incorporates a network-aware scheduling algorithm (based on max flow algorithm) to determine which chunks should be pulled from which site in order to reconstruct the corresponding image on the destination site, with minimal (provisioning) time.

Functional Description: Geo-distributed Storage System to optimize Images (VM, containers, ...) management, in terms of cost and time, in geographically distributed cloud environment (i.e. data centers are connected over WAN).

- Authors: Jad Darrous, Shadi Ibrahim and Christian Pérez
- Contact: Shadi Ibrahim
- URL: https://gitlab.inria.fr/jdarrous/nitro

6.3. VMPlaces

Keywords: Simulation - Virtualization - Scheduling

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, Jonathan Pastor and Mario Südholt
- Contact: Adrien Lèbre
- 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
- URL: http://enos.readthedocs.io/en/stable/

6.5. EnOSlib

EnOSlib is a library to help you with your experiments

Keywords: Distributed Applications - Distributed systems - Evaluation - Grid Computing - Cloud computing - Experimentation - Reproducibility - Linux - Virtualization

FUNCTIONAL DESCRIPTION: EnOSlib is a library to help you with your distributed application experiments. The main parts of your experiment logic is made reusable by the following EnOSlib building blocks:

- Reusable infrastructure configuration: The provider abstraction allows you to run your experiment on different environments (locally with Vagrant, Grid'5000, Chameleon and more) - Reusable software provisioning: In order to configure your nodes, EnOSlib exposes different APIs with different level of expressivity - Reusable experiment facilities: Tasks help you to organize your experimentation workflow.

EnOSlib is designed for experimentation purpose: benchmark in a controlled environment, academic validation . . .

- Contact: Matthieu Simonin
- Publications: Toward a Holistic Framework for Conducting Scientific Evaluations of OpenStack - EnosStack: A LAMP-like stack for the experimenter
- URL: https://discovery.gitlabpages.inria.fr/enoslib/

6.6. Concerto

KEYWORDS: Reconfiguration - Distributed Software - Component models - Dynamic software architecture

FUNCTIONAL DESCRIPTION: Concerto is an implementation of the formal model Concerto written in Python. Concerto aloows to : 1. describe the life-cycle and the dependencies of software components, 2. describe a components assembly that forms the overall life-cycle of a distributed software, 3. automatically reconfigure a Concerto assembly of components by using a set of reconfiguration instructions as well as a formal operational semantics.

- Partners: IMT Atlantique - LS2N - LIP
- Contact: Maverick Chardet
- URL: https://gitlab.inria.fr/VeRDi-project/concerto

6.7. Platforms

6.7.1. OpenStack

OpenStack is the de facto open-source management system to operate and use Cloud Computing infrastructures. Started in 2012, the OpenStack foundation gathers 500 organizations including groups such as Intel, AT&T, RedHat, etc. The software platform relies on tens of services with a 6-month development cycle. It is composed of more than 2 millions of lines of code, mainly in Python, just for the core services. While these aspects make the whole ecosystem quite swift, they are also good signs of maturity of this community. We created and animated between 2016 and 2018 the Fog/Edge/Massively Distributed (FEMDC) Special Interest Group[0] and have been contributing to the Performance working group since 2015. The former investigates how OpenStack can address Fog/Edge Computing use cases whereas the latter addresses scalability, reactivity and high-availability challenges. In addition to releasing white papers and guidelines [96], the major result from the academic view point is the aforementioned EnOS solution, a holistic framework to conduct performance evaluations of OpenStack (control and data plane). In May 2018, the FEMDC SiG turned into a larger group under the control of the OpenStack foundation. This group gathers large companies such as Verizon, ATT, etc.

6.7.2. Grid'5000

Grid'5000 is a large-scale and versatile testbed for experiment-driven research in all areas of computer science, with a focus on parallel and distributed computing including Cloud, HPC and Big Data. It provides access to a large amount of resources: 12000 cores, 800 compute-nodes grouped in homogeneous clusters, and featuring various technologies (GPU, SSD, NVMe, 10G and 25G Ethernet, Infiniband, Omni-Path) and advanced monitoring and measurement features for traces collection of networking and power consumption, providing a deep understanding of experiments. It is highly reconfigurable and controllable. Researchers can

[0] https://wiki.openstack.org/wiki/Fog_Edge_Massively_Distributed_Clouds

experiment with a fully customized software stack thanks to bare-metal deployment features, and can isolate their experiment at the networking layer advanced monitoring and measurement features for traces collection of networking and power consumption, providing a deep understanding of experiments designed to support Open Science and reproducible research, with full traceability of infrastructure and software changes on the testbed. STACK members are strongly involved into the management and the supervision of the testbed, notably through the steering committee or the SeDuCe testbed described hereafter.

6.7.3. SeDuCe

The SeDuCe Project aims to deliver a research testbed dedicated to holistic research studies on energetical aspects of datacenters. Part of the Grid'5000 Nantes' site, this infrastructure is composed of probes that measure the power consumption of each server, each switch and each cooling system, and also measure the temperature at the front and the back of each servers. These sensors enable reasearch to cover a full spectrum of the energetical aspect of datacenters, such as cooling and power consumption depending of experimental conditions.

The testbed is connected to renewable energy sources (solar panels). This "green" datacenter will enable researchers to perform real experiment-driven studies on fields such as temperature based scheduling or "green" aware software (*i.e.*, software that take into account renewable energies and weather conditions).

6.7.4. SILECS

STACK Members are involved in the definition and bootstrap of the SILECS infrastructure. This infrastructure can be seen as a merge of the Grid'5000 and FIT testbeds with the goal of providing a common platform for experimental computer Science (Next Generation Internet, Internet of things, clouds, HPC, big data, etc.).

7. New Results

7.1. Resource Management

Participants: Mohamed Abderrahim, Adwait Jitendra Bauskar, Emile Cadorel, Hélène Coullon, Jad Darrous, David Espinel, Shadi Ibrahim, Thomas Lambert, Adrien Lebre, Jean-Marc Menaud, Alexandre Van Kempen.

In 2019, we achieved several contributions regarding the management of resources and data of cloud infrastructures, especially in a geo-distributed context (*e.g.*, Fog and Edge computing).

The first contributions are related to improvements of low-level building blocks. The following ones deal with geo-distributed considerations. Finally the last ones are related to capacity and placement strategies of distributed applications and scientific workflows.

In [15], we discuss how to improve I/O fairness and SSDs' utilization through the introduction of a NCQ-aware I/O scheduling scheme, NASS. The basic idea of NASS is to elaborately control the request dispatch of workloads to relieve NCQ conflict and improve NCQ utilization at the same time. To do so, NASS builds an evaluation model to quantify important features of the workload. In particular, the model first finds aggressive workloads, which cause NCQ conflict, based on the request size and the number of requests of the workloads. Second, it evaluates merging tendency of each workload, which may affect the bandwidth and cause NCQ conflict indirectly, based on request merging history. Third, the model identifies workloads with deceptive idleness, which cause low NCQ utilization, based on historical requests in I/O scheduler. Then, based on the model, NASS sets the request dispatch of each workload to guarantee fairness and improve device utilization: (1) NASS limits aggressive workloads to relieve NCQ conflict; (2) it adjusts merging of sequential workloads to improve bandwidth of the workloads while relieving NCQ conflict; and (3) it restricts request dispatch of I/O scheduler, rather than stopping request dispatch to improve NCQ utilization. We integrate NASS into four state-of-the-art I/O schedulers including CFQ, BFQ, FlashFQ, and FIOPS. The experimental results show that with NASS, I/O schedulers can achieve 11-23% better fairness and at the same time improve device utilization by 9-29%.

In [16], [28], we address the challenge related to the boot duration of virtual machines and containers in high consolidated cloud scenarios.This time, which can last up to minutes, is critical as it defines how an application can react w.r.t. demands' fluctuations (horizontal elasticity). Our contribution is the YOLO proposal (You Only Load Once). YOLO reduces the number of I/O operations generated during a boot process by relying on a boot image abstraction, a subset of the VM/container image that contains data blocks necessary to complete the boot operation. Whenever a VM or a container is booted, YOLO intercepts all read accesses and serves them directly from the boot image, which has been locally stored on fast access storage devices (e.g., memory, SSD, etc.). In addition to YOLO, we show that another mechanism is required to ensure that files related to VM/container management systems remain in the cache of the host OS. Our results show that the use of these two techniques can speed up the boot duration 2–13 times for VMs and 2 times for containers. The benefit on containers is limited due to internal choices of the docker design. We underline that our proposal can be easily applied to other types of virtualization (e.g., Xen) and containerization because it does not require intrusive modifications on the virtualization/container management system nor the base image structure.

Complementary to the previous contribution and in an attempt to demonstrate the importance of container image placement across edge servers, we propose and evaluate through simulation two novel container image placement algorithms based on k-Center optimization in [14]. In particular, we introduce a formal model to tackle down the problem of reducing the maximum retrieval time of container images, which we denote as MaxImageRetrievalTime. Based on the model, we propose KCBP and KCBP-WC, two placement algorithms which target reducing the maximum retrieval time of container images from any edge server. While KCBP is based on a k-Center solver (i.e., placing k facilities on a set of nodes to minimize the distance from any node to the closet facility) which is applied on each layer and its replicas (taking into account the storage capacities of the nodes), KCBP-WC uses the same principle but it tries to avoid simultaneous downloads from the same node. More precisely, if two layers are part of the same image, then they cannot be placed on the same nodes. We have implemented our proposed algorithms alongside two other state-of-the-art placement algorithms (i.e., Best-Fit and Random) in a simulator written in Python. Simulation results show that the proposed algorithms can outperform state-of- the-art algorithms by a factor of 1.1x to 4x depending on the characteristics of the networks.

In [13], we conduct experiments to thoroughly understand the performance of data-intensive applications under replication and EC. We use representative benchmarks on the Grid'5000 testbed to evaluate how analytic workloads, data persistency, failures, the back-end storage devices, and the network configuration impact their performances. While some of our results follow our intuition, others were unexpected. For example, disk and network contentions caused by chunks distribution and the unawareness of their functionalities are the main factor affecting the performance of data-intensive applications under EC, not data locality. An important outcome of our study is that it illustrates in practice the potential benefits of using EC in data-intensive clusters, not only in reducing the storage cost – which is becoming more critical with the wide adoption of high-speed storage devices and the explosion of generated and to be processed data – but also in improving the performance of data-intensive applications. We extended our work to Fog infrastructures in [31]. In particular, we empirically demonstrate the impact of network heterogeneity on the execution time of MR applications when running in the Fog.

In [5], we propose a first approach to deal with the data location challenges in geo-distribtued object stores. Existing solutions, relying on a distributed hash table to locate the data, are not efficient because location record may be placed far away from the object replicas. In this work, we propose to use a tree-based approach to locate the data, inspired by the Domain Name System (DNS) protocol. In our protocol, servers look for the location of an object by requesting successively their ancestors in a tree built with a modified version of the Dijkstra's algorithm applied to the physical topology. Location records are replicated close to the object replicas to limit the network traffic when requesting an object. We evaluate our approach on the Grid'5000 testbed using micro experiments with simple network topologies and a macro experiment using the topology of the French National Research and Education Network (RENATER). In this macro benchmark, we show that the time to locate an object in our approach is less than 15 ms on average which is around 20% shorter than using a traditional Distributed Hash Table (DHT).

In [20], we present the design, implementation, and evaluation of F-Storm, an FPGA-accelerated and general-purpose distributed stream processing system in the Edge. By analyzing current efforts to enable stream data processing in the Edge and to exploit FPGAs for data-intensive applications, we derive the key design aspects of F-Storm. Specifically, F-Storm is designed to: (1) provide a light-weight integration of FPGA with a DSP system in Edge servers, (2) make full use of FPGA resources when assigning tasks, (3) relieve the high overhead when transferring data between Java Virtual Machine (JVM) and FPGAs, and importantly (4) provide programming interface for users that enable them to leverage FPGA accelerators easily while developing their stream data applications. We have implemented F-Storm based on Storm. Evaluation results show that F-Storm reduces the latency by 36% and 75% for matrix multiplication and grep application compared to Storm. Furthermore, F-Storm obtains 1.4x, 2.1x, and 3.4x throughput improvement for matrix multiplication, grep application, and vector addition, respectively.

In [30], we discuss the main challengres related to the design and development of inter-site services for operating a massively distributed Cloud-Edge architecture deployed in different locations of the Internet backbone (i.e, network point of presences). More precisely, we discuss challenges related to the establishment of connectivity among several virtual infrastructure managers in charge of operating each site. Our goal is to initiate the discussion about the research directions on this field providing some interesting points to promote future work.

In [7], we focus on how to reduce the costly cross-rack data transferring in MapReduce systems. We observe that with high Map locality, the network is mainly saturated in Shuffling but relatively free in the Map phase. A little sacrifice in Map locality may greatly accelerate Shuffling. Based on this, we propose a novel scheme called Shadow for Shuffle-constrained general applications, which strikes a trade-off between Map locality and Shuffling load balance. Specifically, Shadow iteratively chooses an original Map task from the most heavily loaded rack and creates a duplicated task for it on the most lightly loaded rack. During processing, Shadow makes a choice between an original task and its replica by efficiently pre-estimating the job execution time. We conduct extensive experiments to evaluate the Shadow design. Results show that Shadow greatly reduces the cross-rack skewness by 36.6% and the job execution time by 26% compared to existing schemes.

In [6], we consider a complete framework for straggler detection and mitigation. We start with a set of metrics that can be used to characterize and detect stragglers including Precision, Recall, Detection Latency, Undetected Time and Fake Positive. We then develop an architectural model by which these metrics can be linked to measures of performance including execution time and system energy overheads. We further conduct a series of experiments to demonstrate which metrics and approaches are more effective in detecting stragglers and are also predictive of effectiveness in terms of performance and energy efficiencies. For example, our results indicate that the default Hadoop straggler detector could be made more effective. In certain case, Precision is low and only 55% of those detected are actual stragglers and the Recall, i.e., percent of actual detected stragglers, is also relatively low at 56%. For the same case, the hierarchical approach (i.e., a green-driven detector based on the default one) achieves a Precision of 99% and a Recall of 29%. This increase in Precision can be translated to achieve lower execution time and energy consumption, and thus higher performance and energy efficiency; compared to the default Hadoop mechanism, the energy consumption is reduced by almost 31%. These results demonstrate how our framework can offer useful insights and be applied in practical settings to characterize and design new straggler detection mechanisms for MapReduce systems.

In [21], we provide a general solution for workflow performance optimizations considering system variations. Specifically, we model system variations as time-dependent random variables and take their probability distributions as optimization input. Despite its effectiveness, this solution involves heavy computation overhead. Thus, we propose three pruning techniques to simplify workflow structure and reduce the probability evaluation overhead. We implement our techniques in a runtime library, which allows users to incorporate efficient probabilistic optimization into existing resource provisioning methods. Experiments show that probabilistic solutions can improve the performance by 51% compared to state-of-the-art static solutions while guaranteeing budget constraint, and our pruning techniques can greatly reduce the overhead of probabilistic optimization.

In [11], we propose a new strategy to schedule heteregeneous scientific workflows while minimizing the energy consumption of the cloud provider by introducing a deadline sensitive algorithm. Scheduling workflows in a cloud environment is a difficult optimization problem as capacity constraints must be fulfilled additionally to dependencies constraints between tasks of the workflows. Usually, work around the scheduling of scientific workflows focuses on public clouds where infrastructure management is an unknown black box. Thus, many works offer scheduling algorithms designed to select the best set of virtual machines over time, so that the cost to the end user is minimized. This paper presents the new v-HEFT-*deadline* algorithm that takes into account users deadlines to minimize the number of machines used by the cloud provider. The results show the real benefits of using our algorithm for reducing the energy consumption of the cloud provider.

In [9],we investigate how a monitoring service for Edge infrastructures should be designed in order to mitigate as much as possible its footprint in terms of used resources. Monitoring functions tend to become compute-, storage-and network-intensive, in particular because they will be used by a large part of applications that rely on real-time data. To reduce as much as possible the footprint of the whole monitoring service, we propose to mutualize identical processing functions among different tenants while ensuring their quality-of-service (QoS) expectations. We formalize our approach as a constraint satisfaction problem and show through micro-benchmarks its relevance to mitigate compute and network footprints.

In [22], we propose a generalization of the previous work. More precisely, weinvestigates whether the use of Constraint Programming (CP) could enable the development of a generic and easy-to-upgrade placement service for Fog/Edge Computing infrastructures. Our contribution is a new formulation of the placement problem, an implementation of this model leveraging Choco-solver and an evaluation of its scalability in comparison to recent placement algorithms. To the best of our knowledge, our study is the first one to evaluate the relevance of CP approaches in comparison to heuristic ones in this context. CP interleaves inference and systematic exploration to search for solutions, letting users on what matters: the problem description. Thus, our service placement model not only can be easily enhanced (deployment constraints/objectives) but also shows a competitive tradeoff between resolution times and solutions quality.

In [27], we present the first building blocks of a simulator to investigate placement challenges in Edge infrastructures. Efficiently scheduling computational jobs with data-sets dependencies is one of the most important challenges of fog/edge computing infrastructures. Although several strategies have been proposed, they have been evaluated through ad-hoc simulator extensions that are, when available, usually not maintained. This is a critical problem because it prevents researchers to-easily-conduct fair evaluations to compare each proposal. We propose to address this limitation throught the design and development of a common simulator. More precisely, in this research report, we describe an ongoing project involving academics and a high-tech company that aims at delivering a dedicated tool to evaluate scheduling policies in edge computing infrastructures. This tool enables the community to simulate various policies and to easily customize researchers/engineers' use-cases, adding new functionalities if needed. The implementation has been built upon the Batsim/SimGrid toolkit, which has been designed to evaluate batch scheduling strategies in various distributed infrastructures. Although the complete validation of the simulation toolkit is still ongoing, we demonstrate its relevance by studying different scheduling strategies on top of a simulated version of the Qarnot Computing platform, a production edge infrastructure based on smart heaters.

In [8], we propose an efficient graph partitioning method named Geo-Cut, which takes both the cost and performance objectives into consideration for large graph processing in geo-distributed DCs.Geo-Cut adopts two optimization stages. First, we propose a cost-aware streaming heuristic and utilize the one-pass streaming graph partitioning method to quickly assign edges to different DCs while minimizing inter-DC data communication cost. Second, we propose two partition refinement heuristics which identify the performance bottlenecks of geo-distributed graph processing and refine the partitioning result obtained in the first stage to reduce the inter-DC data transfer time while satisfying the budget constraint. Geo-Cut can be also applied to partition dynamic graphs thanks to its lightweight runtime overhead. We evaluate the effectiveness and efficiency of Geo-Cut using real-world graphs with both real geo-distributed DCs and simulations. Evaluation results show that Geo-Cut can reduce the inter-DC data transfer time by up to 79% (42% as the median) and reduce the monetary cost by up to 75% (26% as the median) compared to state-of-the-art graph partitioning methods with a low overhead.

7.2. Programming Support

Participants: Maverick Chardet, Hélène Coullon, Thomas Ledoux, Jacques Noyé, Dimitri Pertin, Simon Robillard, Hamza Sahli, Charlène Servantie.

Our contributions regarding programming support are divided in two topics. First, we focused on one specific challenge related to distributed software deployment: distributed software commissioning. We have proposed a useful approach for introducing model checking to help system operators design their parallel distributed software commissioning. Then, we focused on Fog formalization and we have proposed a fully graphical process algebraic formalism to design a Fog system.

In [12], MADA, a deployment approach to facilitate the design of efficient and safe distributed software commissioning is presented. MADA is built on top of the Madeus formal model that focuses on the efficient execution of installation procedures. Madeus puts forward more parallelism than other commissioning models, which implies a greater complexity and a greater propensity for errors. MADA provides a new specific language on top of Madeus that allows the developer to easily define the properties that should be ensured during the commissioning process. Then, MADA automatically translates the description to a time Petri net and a set of TCTL formulae. MADA is evaluated on the OpenStack commissioning.

About Fog formalization, we present a novel formal model defining spatial and structural aspects of Fog-based systems using Bigraphical Reactive Systems, a fully graphical process algebraic formalism [17]. The model is extended with reaction rules to represent the dynamic behavior of Fog systems in terms of self-adaptation. The notion of bigraph patterns is used in conjunction with boolean and temporal operators to encode spatio-temporal properties inherent to Fog systems and applications. The feasibility of the modelling approach is demonstrated via a motivating case study and various self-adaptation scenarios.

Overall, the number of contributions we made this year on the programming support topic is less significative than the previous one. However, we would like to underline that it does not reflect the recent efforts we put. In particular, the team has strongly developed the field of dynamic reconfiguration of distributed software systems and expects to get important results during 2020.

7.3. Energy-aware computing

Participants: Emile Cadorel, Hélène Coullon, Adrien Lebre, Thomas Ledoux, Jean-Marc Menaud, Jonathan Pastor, Dimitri Saingre, Yewan Wang.

Energy consumption is one of the major challenges of modern datacenters and supercomputers. Our works in Energy-aware computing can be categorized into two subdomains: Software level (SaaS, PaaS) and Infrastructure level (IaaS). At Software level, we worked on the general Cloud applications architectures and more recently on BlockChain-based solutions. At Infrastructure level, we worked this year on two directions: (i) investigating the thermal aspects in datacenters, and (ii) analyzing the energy footprint of geo-distributed plateforms.

In [11], the scheduling of heterogeneous scientific workflows while minimizing the energy consumption of the cloud provider is tackled by introducing a deadline sensitive algorithm. Scheduling workflows in a cloud environment is a difficult optimization problem as capacity constraints must be fulfilled additionally to dependencies constraints between tasks of the workflows. Usually, work around the scheduling of scientific workflows focuses on public clouds where infrastructure management is an unknown black box. Thus, many works offer scheduling algorithms designed to select the best set of virtual machines over time, so that the cost to the end user is minimized. This paper presents the new v-HEFT-*deadline* algorithm that takes into account users deadlines to minimize the number of machines used by the cloud provider. The results show the real benefits of using our algorithm for reducing the energy consumption of the cloud provider.

In [25], over the last year, both academic and industry have increase their work on blockchain technologies. Despite the potential of blockchain technologies in many areas, several obstacles are slowing down their development. In addition to the legal and social obstacles, technical limitations now prevent them from imposing themselves as a real alternative to centralised services. For example, several problems dealing with the scalability or the energy cost have been identified. That's why, a significant part of this research is focused on improving the performances (latency, throughput, energy footprint, etc.) of such systems. Unfortunately, Those projects are often evaluated with ad hoc tools and experimental environment, preventing reproducibility and easy comparison of new contribution to the state of the art. As a result, we notice a clear lack of tooling concerning the benchmarking of blockchain technologies. To the best of our knowledge only a few tools address such issues. Those tools often relies on the load generation aspect and omit some other important aspect of benchmark experiments such as reproducibility and the network emulation. We introduce BCTMark, a general framework for benchmarking blockchain technologies in an emulated environment in a reproductible way.

In [18], we present a deep evaluation about the power models based on CPU utilization. The influence of inlet temperature on models has been especially discussed. According to the analysis, one regression formula by using CPU utilization as the only indicator is not adequate for building reliable power models. First of all, Workloads have different behaviors by using CPU and other hardware resources in server platforms. Therefore, power is observed to have high dispersion for a fixed CPU utilization, especially at full workload. At the same time, we also find that, power is well proportional to CPU utilization within the execution of one single workload. Hence, applying workload classifications could be an effective way to improve model accuracy. Moreover, inlet temperature can cause surprising influence on model accuracy. The model reliability can be questioned without including inlet temperature data. In a use case, after including inlet temperature data, we have greatly improved the precision of model outputs while stressing server under three different ambient temperatures.

In [18], our physical experiments have shown that even under the same conditions, identical processors consume different amount of energy to complete the same task. While this manufacturing variability has been observed and studied before, there is lack of evidence supporting the hypotheses due to limited sampling data, especially from the thermal characteristics. In this article, we compare the power consumption among identical processors for two Intel processors series with the same TDP (Thermal Design Power) but from different generations. The observed power variation of the processors in newer generation is much greater than the older one. Then, we propose our hypotheses for the underlying causes and validate them under precisely controlled environmental conditions. The experimental results show that, with the increase of transistor densities, difference of thermal characteristics becomes larger among processors, which has non-negligible contribution to the variation of power consumption for modern processors. This observation reminds us of re-calibrating the precision of the current energy predictive models. The manufacturing variability has to be considered when building energy predictive models for homogeneous clusters.

In [3], we propose a model and a first implementation of a simulator in order to compare the energy footprint of different cloud architectures (single sites vs fully decentrlaized). Despite the growing popularity of Fog/Edge architectures, their energy consumption has not been well investigated yet. To move forward on such a critical question, we first introduce a taxonomy of different Cloud-related architectures. From this taxonomy, we then present an energy model to evaluate their consumption. Unlike previous proposals, our model comprises the full energy consumption of the computing facilities, including cooling systems, and the energy consumption of network devices linking end users to Cloud resources. Finally, we instantiate our model on different Cloud-related architectures, ranging from fully centralized to completely distributed ones, and compare their energy consumption. The results validates that a completely distributed architecture, because of not using intra-data center network and large-size cooling systems, consumes less energy than fully centralized and partly distributed architectures respectively. To the best of our knowledge, our work is the first one to propose a model that enables researchers to analyze and compare energy consumption of different Cloud-related architectures.

7.4. Security and Privacy

Participants: Mohammad-Mahdi Bazm, Fatima Zahra Boujdad, Wilmer Edicson Garzon Alfonso, Jean-Marc Menaud, Sirine Sayadi, Mario Südholt.

This year the team has provided two major contributions on security and privacy challenges in distributed systems. First, we have extended our model for secure and privacy-aware biomedical analyses, as well as started to explore the impact of the big-data analyses in this context. Second, we have contributed mitigation methods for Cloud-based side-channel attacks.

In [24], we have developed a methodology for the development of secure and privacy-aware biomedical analyses we motivate the need for real distributed biomedical analyses in the context of several ongoing projects, including the I-CAN project that involves 34 French hospitals and affiliated research groups. We present a set of distributed architectures for such analyses that we have derived from discussions with different medical research groups and a study of related work. These architectures allow for scalability, security/privacy and reproducibility properties to be taken into account. A predefined set of architectures allows medecins and biomedical engineers to define high-level distributed architectures for biomedical analyses that ensure strong security and constraints on private data. Architectures from this set can then be implemented with ease because of detailed, also predefined, detailed implementation templates. Finally, we illustrate how these architectures can serve as the basis of a development method for biomedical distributed analyses.

In [10] and [23], we presented a new taxonomy for container security with a particular focus on data transmitted through the virtualization boundary. Containerization is a lightweight virtualization technique reducing virtualization overhead and deployment latency compared to full VM; its popularity is quickly increasing. However, due to kernel sharing, containers provide less isolation than full VM. Thus, a compromised container may break out of its isolated context and gain root access to the host server. This is a huge concern, especially in multi-tenant cloud environments where we can find running on a single server containers serving very different purposes, such as banking microservices, compute nodes or honeypots. Thus, containers with specific security needs should be able to tune their own security level. Because OS-level defense approaches inherited from time-sharing OS generally requires administrator rights and aim to protect the entire system, they are not fully suitable to protect usermode containers. Research recently made several contributions to deliver enhanced security to containers from host OS level to (partially) solve these challenges. In this survey, we propose a new taxonomy on container defense at the infrastructure level with a particular focus on the virtualization boundary, where interactions between kernel and containers take place. We then classify the most promising defense frameworks into these categories.

Finally, we have leveraged an approach based on Moving Target Defense (MTD) theory to interrupt a cache-based side-channel attack between two Linux containers in the context of the Mohammad Mahdi's PhD thesis [1]. MTD allows us to make the configuration of system more dynamic and consequently more harder to attack by an adversary, by using shuffling at different level of system and cloud. Our approach does not need to carrying modification neither into the guest OS or the hypervisor. Experimental results show that our approach imposes very low performance overhead. We have also provided a survey on the isolation challenge and on the cache-based side-channel attacks in cloud computing infrastructures. We have developed different approaches to detect/mitigate cross-VM/cross-containers cache-based side-channel attacks. Regarding the detection of cache-based side-channel attacks, we have enabled their detection by leveraging Hardware performance Counters (HPCs) and Intel Cache Monitoring Technology (CMT) with anomaly detection approaches to identify a malicious virtual machine or a Linux container. Our experimental results show a high detection rate.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

Participants: Ronan-Alexandre Cherrueau, Marie Delavergne, Adrien Lebre [Contact point], Javier Rojas Balderrama, Matthieu Simonin.

Following the ENOS bilateral contract ("Contrat de Recherche Externalisé") between Orange and Inria (Sept 2017-Oct 2018), we agreed with Orange Labs to pursue this collaboration around a second contrat. This new contrat, which is going to last 18 months for a budget of 150K€, targets the following objectives:

- Strengthen the Enos framework and the resulting EnosLib solution (see Section 6.4 and Section 6.5).
- Define an experimental protocol allowing the automatozed and reproducible evaluation of an OpenStack instance in a WANWide context.
- Develop a DSL to reify location aspects at the CLI level in order to create new resources (image, VM, etc.) through a set of OpenStack instances while guaranteeing a notion of master copy.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. SysMics

Participants: Jean-Marc Menaud, Mario Südholt [coordinator].

The SysMics project aims at federating the NExT scientific community toward a common objective: anticipate the emergence of systems medicine by co-developing 3 approaches in population-scale genomics: genotyping by sequencing, cell-by-cell profiling and microbiome analysis. STACK investigates new means for secure and privacy-aware computations in the context of personalized medecine, notably genetic analyses.

This project is financed by the Nantes excellency initiative in Medecine and Informatics (NExT) from 2018-22.

9.1.2. SHLARC

Participants: Mario Südholt [coordinator], Sirine Sayadi.

The SHLARC project is an international network involving more than 20 partners from more than 15 countries located on four continents. The network aims at improving HLA imputation techniques in the domain of immunobiology, notably by investigation better computational methods for the correspoding biomedical analyses.

The ambition of the SHLARC is to bring together international expertise to solve essential questions on immune-related pathologies through innovative algorithms and powerful computation tool development. To achieve this goal, we determined 3 main objectives

- Data. By bringing together scientists from around the world, we will collectively increase the amount of SNP+HLA data available, both in terms of quantity and diversity.
- Applied mathematical and computer sciences. We will further optimize SNP-HLA imputation methods using the attribute-bagging HIBAG tool, and particularly for genetically diverse and admixed populations.
- Accessibility and service to the scientific community. Following the Haplotype Reference Consortium (HRC) initiative, the network envisions building a free, user-friendly webserver where researchers can access improved imputation protocols by simply uploading their data and obtaining the best possible HLA imputation for their dataset.

In this context, the STACK team is working on improved analysis techniques that harness distributed infrastructures.

This project is financed by the Nantes excellency initiative in Medecine and Informatics (NExT) from 2019-22.

9.1.3. Oncoshare

Participant: Mario Südholt [coordinator].

The ONCOSHARe project (ONCOlogy big data SHAring for Research) will demonstrate, through a multi-disciplinary cooperation within the Western CANCEROPOLE network, the feasibility and the added value of a Cancer Patient Centered Information Common for in-silico research. The STACK team will work on challenges to the security and the privacy of user data in this context.

This project is financed by three French regions from 2018-2021.

9.2. National Initiatives

9.2.1. Ademe

9.2.1.1. GL4MA

Participants: Brice Nédelec, Thomas Ledoux [coordinator].

The Green Label for Microservices Architecture (GL4MA) project aims to design and develop a technological platform (tools, framework, dedicated languages) for the self management of eco-responsible micro-service architectures for the Cloud. The experiments will be carried out through case studies provided by Sigma Informatique and the presence of renewable energy will initially be simulated. At the end of the project, the technological platform will be deployed as part of the CPER SeDuCe platform. This project is founded by the Ademe (call Perfecto) running for 18 months (starting in September 2019).

9.2.2. CominLabs laboratory of excellence

9.2.2.1. 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 STACK team. The project provides funding of 330 KEUR altogether with an STACK 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 STACK team is mainly involved in providing solutions for the second and third challenges.

9.2.2.2. SeDuCe++

Participants: Jonathan Pastor, Jean-Marc Menaud [coordinator].

SeDuCe++ is an extended version of the SeDuCe project. Funded by the LS2N (CNRS), an allocated budget of 10K€ for one year, it aims at studying the energy footprint of extreme edge infrastructure.

9.2.3. ANR

9.2.3.1. GRECO (ANR)

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

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 522K€, 90K€ for STACK).

The consortium is composed of 4 partners: Qarnot Computing (coordinator) and 3 academic research group (DATAMOVE and AMA from the LIG in Grenoble and STACK 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. To move forward to this objective, we have been designing a simulator to innovate in designing scheduling and data management systems. This simular leverage the Simgrid/PyBATSIM solution [27].

9.2.3.2. 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 238K€).

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.

9.2.4. FSN

9.2.4.1. Hydda (FSN)

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

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 questions we are investigating 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?

9.2.5. CPER

9.2.5.1. SeDuCe

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

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).

9.2.6. Inria Project Labs

9.2.6.1. DISCOVERY

Participants: Javier Rojas Balderrama, Hélène Coullon, Marie Delavergne, Shadi Ibrahim, Adrien Lebre [coordinator], Ronan-Alexandre Cherrueau, Matthieu Simonin, Alexandre Van Kempen.

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 [0] 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.

STACK led the DISCOVERY IPL and contributes mainly around two axes: VM life cycle management and deployment/reconfiguration challenges.

The IPL ended in July 2019.

9.2.7. InriaHub

9.2.7.1. Mercury

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

STACK, 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.

The Mercury action ended in July 2019.

9.2.7.2. Apollo/Soyuz

Participants: Javier Rojas Balderrama, Ronan-Alexandre Cherrueau, Adrien Lebre [coordinator], Matthieu Simonin.

The Apollo/Soyuz is the second InriaHub action attached the DISCOVERY IPL. While Mercury aims mainly at supporting development efforts within the DISCOVERY IPL, the Apollo/Soyuz is focusing on the animation and the dissemination of the DISCOVERY activities within the different open-source ecosystem (*i.e.*, OpenStack, OPNFV, etc.). One additional engineer will join the current team in January 2019.
Further information available at: http://beyondtheClouds.github.io.

The Apollo/Soyuz ended in Dec 2019.

9.2.8. Fonds d'amorçage IMT Industrie du Futur 2017

9.2.8.1. aLIFE

Participants: Hélène Coullon [coordinator], Jacques Noyé.

As a follow-up of the aLIFE workshop (Nantes, Jan. 2018), organized in partnership with colleagues from IMT Atlantique and gathering both academic and industrial partners, we have written a booklet [29] summarizing the workshop discussions and proposing a shared vision of what software research could bring to Industry 4.0 initiatives.

[0] http://beyondtheclouds.github.io

9.2.9. Connect Talent

9.2.9.1. Apollo (Connect Talent)

Participant: Shadi Ibrahim [Coordinator].

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

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.

9.2.10. Etoiles Montantes

9.2.10.1. VeRDi

Participant: Hélène Coullon [Coordinator].

VeRDi is an acronym for Verified Reconfiguration Driven by execution. The VeRDi project is funded by the French region Pays De La Loire where Nantes is located. The project starts in November 2018 and ends on December 2020 with an allocated budget of 172800€.

It aims at addressing distributed software reconfiguration in an efficient and verified way. The aim of the VeRDi project is to build an argued disruptive view of the problem. To do so we want to validate the work already performed on the deployment in the team and extend it to reconfiguration.

9.3. International Initiatives

9.3.1. Inria International Labs

Inria@SiliconValley

Associate Team involved in the International Lab:

9.3.1.1. Hermes

Title: Accelerating the Performance of Multi-Site Scientific applications through Coordinated Data management.

International Partner (Institution - Laboratory - Researcher):

Lawrence Berkeley National Laboratory (United States) - Scientific Data Management Group - Suren Byna.

Start year: 2019

See also: http://hermes-ea2019.gforge.inria.fr.

Advances in computing, experimental, and observational facilities are enabling scientists to generate and analyze unprecedented volumes of data. A critical challenge facing scientists in this era of data deluge is storing, moving, sharing, retrieving, and gaining insight from massive collections of data efficiently. Existing data management and I/O solutions on high-performance computing (HPC) systems require significant enhancements to handle the three V's of Big Data (volume, velocity, and variety) in order to improve productivity of scientists. Even more challenging, many scientific Big Data and machine learning applications require data to be shared, exchanged, and transferred among multiple HPC sites. Towards overcoming these challenges, in this project, we aim at accelerating scientific Big Data application performance through coordinated data management that addresses performance limitations of managing data across multiple sites. In particular, we focus on challenges related to the management of data and metadata across sites, distributed burst buffers, and online data analysis across sites.

9.3.2. Inria International Partners

9.3.2.1. Informal International Partners

Huazhong university of Science and Technology (HUST): We collaborate on resource management for stream data applications in the edge, I/O scheduling for SDDs and network-aware task scheduling for MapReduce.

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

ShenZhen University: We collaborate on resource management for workflows in the clouds and optimizing graph processing in geo-distributed data-centers.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Suren Byna, a Staff Scientist in the Scientific Data Management Group at Lawrence Berkeley National Lab (LBNL), visited the STACK team from September 30 to October 4 2019. This visit was in the context of the Hermes Associate team.
- Twinkle Jain, a PhD student at Northeastern university, visited the STACK team from May 1 to July 31 2019. Twinkle was working with S. Ibrahim on stragglers mitigation in big data systems. The visit was funded by the ANR KerStream and the Apollo Connect Talent projects.

9.4.1.1. Internships

- Asha Begam Mohamed Mubarak, a master student at University of Rennes 1, joined the team as a research intern from April 2019 until August 2019. Her thesis was on fast Container Image Retrieval in the Edge.

9.4.2. Visits to International Teams

9.4.2.1. Research Stays Abroad

HUST, China: From August 23 to September 2, S. Ibrahim visited the Services Computing Technology and System Lab at Huazhong university of Science and Technology.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

- A. Lebre co-organized the TILECS workshop (Towards an Infrastructure for Large-Scale Experimental Computer Science), Grenoble, July 2019.

10.1.2. Scientific Events: Selection

10.1.2.1. Chair of Conference Program Committees

- S. Ibrahim was the program chair of the IEEE BigDataSE 2019.
- S. Ibrahim was the program chair of the Smart Data-2019.
- S. Ibrahim was the track co-chair for the Cloud Computing Track of CCGrid 2019.

10.1.2.2. Member of the Conference Program Committees

- H. Coullon was a member of the program committees of CCGrid 2019 and ICCS 2019.
- S. Ibrahim was a member of the program committees of SC'19, Poster track of SC 2019, ICPP 2019, Cluster 2019, ICA3PP 2019, PDSW@SC'19, and HPBDC@IPDPS'19.
- A. Lebre was a member of the program comittees of ICC 2019, IC2E 2019, CCGRID 2019, Cluster 2019, and CloudCom 2019.

- T. Ledoux was a member of the program committees of the conference Compas'19, of the 1st Workshop on Service Quality and Quantitative Evaluation in new Emerging Technologies @ IC2E'19, CrossCloud'19@CCGrid, and ARM'19@Middleware.
- J.-M. Menaud was a member of the program committees of SMARTGREENS'19, SOFTCOM 19.
- M. Südholt was a member of the program committees of CCGrid'19 and CloudCom'19.

10.1.2.3. Member of the Conference Steering Committees

- A. Lebre is a member of the steering committee of the international conference of Fog and Edge Computing (ICFEC).
- M. Südholt is a member of the steering committee of the international conference Programming.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- A. Lebre is an Associate Editor of the IEEE Transactions on BigData.
- A. Lebre is an Associate Editor of the IEEE Transactions on Cloud Computing.
- S. Ibrahim is an Associate Editor of Springer Frontiers of Computer Science journal.
- M. Südholt is an Associate Editor of the journal Programming.

10.1.3.2. Reviewer - Reviewing Activities

- S. Ibrahim has been a reviewer for the following journals: Future Generation Computer Systems, IEEE Internet Computing.
- A. Lebre has been a reviewer for the IEEE Transactions on Cloud Computing journal.
- T. Ledoux has been a reviewer for the following journals: Journal of Systems Architecture - Elsevier and Future Generation Computer Systems.
- M. Südholt has been a reviewer for the journal Transactions of Software Engineering (IEEE TSE).

10.1.4. Invited Talks

- H. Coullon has given a talk to the WIDE Inria team, "Distributed Software Management: Efficiency, Software Engineering and Verification", Rennes, France, Aug. 2019.
- H. Coullon has given a talk to the joint LaMHa/LTP working group of the GDR GPL, "Efficient and Safe Distributed Software Commissioning", Paris, France, Nov. 2019.
- H. Coullon has given a talk to the Department of Computer Science at the University of Tromsø, the Arctic University of Norway, "Efficient and Safe Distributed Software Commissioning and Reconfiguration", Norway, Dec. 2019.
- A. Lebre has given a talk to the Inria Business Club in Paris, France, May 2019.
- A. Lebre has given a talk to the TILECS Workshop in Grenoble, France, July 2019.
- A. Lebre has given a talk to the Journées Clouds in Toulouse, France, Sept 2019.
- A. Lebre has given a talk to the Inria/Interdigital Workshop in Rennes, France, Dec. 2019.

10.1.5. Scientific Expertise

- A. Lebre is a member of the scientific committe of the joint lab between Inria and Nokia Bell Labs.

10.1.6. Research Administration

- A. Lebre is a 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 a member of the executive and architect committees of the Grid'5000 GIS (Groupement d'intérêt scientifique).
- A. Lebre is a member of the executive committee of the <I/O> Lab, a joint lab between Inria and Orange Labs.

- J. Noyé is the Deputy Head of the Automation, Production and Computer Sciences department of IMT Atlantique.
- J.-M. Menaud is organizer of "Pôle Science du Logiciel et des Systèmes Distribués" in Laboratoire des Sciences du Numérique à Nantes (LS2N). He is involved in the GIS VITTORIA (VIrTual inTegrative Oncology Research and InnovAtion).
- J.-M. Menaud is involved in the GIS PERLE (Pôle d'Excellence de la Recherche Ligérienne en Energie).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

- S. Ibrahim is the co-coordinator of the international Master's program in Cloud Computing and Services at University of Rennes 1.
- T. Ledoux is the head of the apprenticeship program in Software Engineering FIL (http://www.imt-atlantique.fr/formation/ingenieur-par-apprentissage/ingenieurs-specialite-ingenierie-logicielle). This 3-year program leads to the award of a Master degree in Software Engineering from the IMT Atlantique.
- H. Coullon is responsible for the Computer Science domain of the new apprenticeship program in Industry 4.0 (FIT) of IMT Atlantique. This 3-year program leads to the award of a Master degree in Industry 4.0 from the IMT Atlantique.

10.2.2. Supervision

- PhD: Mohammad Mahdi Bazm, co-director: J.M. Menaud, director: M. Südholt, "Architecture d'isolation unifiée et mécanismes de lutte contre les canaux auxiliaires pour infrastructures cloud décentralisées", Defended July 2019, the8th.
- Phd: Linh Thuy Nguyen, director: A. Lebre, "Fast delivery of Virtual Machines and Containers: Understanding and optimizing the boot operation", defended on Sept 2019, the 24th.
- PhD: Jad Darrous, advisor: S. Ibrahim, director: C. Perez (Avalon), "Scalable and Efficient Data Management in Distributed Clouds: Service Provisioning and Data Processing" Defended Dec 2019, the 17th.
- PhD: Dimitri Saingre, advisor: T. Ledoux, director: J-M. Menaud.
- PhD: Maverick Chardet, advisor: H. Coullon, director: C. Perez (Avalon).
- PhD: Emile Cadorel, advisor: H. Coullon, director: J-M. Menaud.
- PhD: Jolan Philippe, adviros: H. Coullon, M. Tisi (NaoMod), director: G. Sunye (NaoMod).
- PhD: Yewan Wang, advisor: J.-M. Menaud.
- PhD: Maxime Belair, advisor: J.-M. Menaud.
- PhD: Fatima-zahra Boujdad, advisor: M. Südholt.
- PhD: Wilmer Garzon, advisor: M. Südholt.
- PhD: Sirine Sayadi, advisor: M. Südholt.
- Postdoc: David Guyon, advisor: S. Ibrahim.
- Postdoc: Thomas Lambert, advisor: S. Ibrahim.
- Postdoc: Jonathan Pastor, advisor: J.-M. Menaud.
- Postdoc: Rémy Pottier, advisor: J.-M. Menaud.
- Postdoc: Simon Robillard, advisor: H. Coullon.
- Postdoc: Hamza Sahli, advisor: T. Ledoux.
- Postdoc: Alexandre Van Kempen: A. Lebre (until June 2018).

- Engineer: Ronan-Alexandre Cherrueau, advisor: A. Lebre.
- Engineer: Brice Nédelec, advisor: T. Ledoux.
- Engineer: Dimitri Pertin, advisor: H. Coullon (Sept, Oct 2019).
- Engineer: Javier Rojas Balderrama, advisor: M. Simonin/A. Lebre.
- Engineer: Charlène Servantie, advisor: H. Coullon.

10.2.3. Juries

- H. Coullon was a reviewer of the Master thesis committee of Jolan Philippe, "Systematic development of efficient programs on parallel data structures", Northern Arizona University USA, Feb. 05, 2019.
- H. Coullon was a reviewer of the PhD committee of Alexandre Da Silva Veith, "Quality of Service Aware Mechanisms for (Re)Configuring Data Stream Processing Applications on Highly Distributed Infrastructure", Univ. Lyon, Sept. 32, 2019.
- H. Coullon was a reviewer of the PhD committee of KsanderEJJAAOUANI, "Conception du modèle de programmation INKS pour la séparation des préoccupations algorithmiques et d'optimisation dans les codes de simulation numérique ; application à la résolution du système Vlasov/Poisson 6D", Univ. Stasbourg, Nov. 25, 2019.
- S. Ibrahim was a member of the PhD committee of Chaopeng GUO, "Energy-efficient Resource Provisioning for Cloud Databases", Université Paul Sabatier, Jun. 14, 2019.
- A. Lebre was a member of the PhD Committee of Ali Reza Zamani Zadeh Najari, "Scheduling Edge And In-Transit Computing Resources For Stream Processing Applications", University of Rutgers (USA), May 2019.
- A. Lebre was a member of the PhD Committee of Luke Bertot "Improving the simulation of IaaS Clouds", University of Strasbourg (France), June 2019.
- A. Lebre was a reviewer and member of the PhD Committee of Loic Perennou, "Virtual Machine Experience Design: A Predictive Resource Allocation Approach for Cloud Infrastructures". Conservatoire National des Arts et Metier Paris (France), Oct 2019.
- A. Lebre was a member of the PhD Committee of Genc Tato "Lazy and Locality-Aware Building blocks for Fog Middleware: A Service Discovery Use Case", University of Rennes (France), Dec 2019.
- T. Ledoux was a member of the PhD committee of Xinxiu Tao, "Reliability of changes in cloud environment at PaaS level", Univ. Grenoble Alpes, Jan. 29, 2019.
- T. Ledoux was a member of the PhD committee of Amir Teshome Wonjiga, "User-centric security monitoring in cloud environments", Univ. Rennes 1, Jun. 03, 2019.
- T. Ledoux was a reviewer of the PhD committee of Umar Ozeer, "Autonomic Resilience of Distributed IoT Applications in the Fog", Univ. Grenoble Alpes, Dec. 11, 2019.
- J.-M. Menaud was a member of the PhD committee of Léo Grange, "Gestion de centre de données avec sources d'énergies renouvelables intermittentes et incertaines", Univ. Toulouse, Oct. 3, 2019.

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

- S. Ibrahim, Online publication http://emergences.inria.fr/2019/newsletter-n59/l59-Kerstream: *Un intergiciel post-Hadoop pour gérer les flux de données*.

10.3.2. Articles and contents

As a follow-up of the aLIFE workshop (Nantes, Jan. 2018), organized in partnership with colleagues from IMT Atlantique and gathering both academic and industrial partners, Hélène Coullon and Jacques Noyé have participated to a booklet publication [29] summarizing the workshop discussions and proposing a shared vision of what software research could bring to Industry 4.0 initiatives.

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] M.-M. BAZM.*Unified isolation architectures and mechanisms against side channel attacks for decentralized cloud infrastructures*, Université de Nantes (UNAM), July 2019, https://hal.inria.fr/tel-02417362

[2] T. L. NGUYEN.*Fast delivery of virtual machines and containers : understanding and optimizing the boot operation*, Ecole nationale supérieure Mines-Télécom Atlantique, September 2019, https://tel.archives-ouvertes.fr/tel-02418752

Articles in International Peer-Reviewed Journal

[3] E. AHVAR, A.-C. ORGERIE, A. LEBRE.*Estimating Energy Consumption of Cloud, Fog and Edge Computing Infrastructures*, in "IEEE Transactions on Sustainable Computing", April 2019, p. 1-12 [*DOI :* 10.1109/TSUSC.2019.2905900], https://hal.archives-ouvertes.fr/hal-02083080

[4] O. BEAUMONT, T. LAMBERT, L. MARCHAL, B. THOMAS.*Performance Analysis and Optimality Results for Data-Locality Aware Tasks Scheduling with Replicated Inputs*, in "Future Generation Computer Systems", October 2019, p. 1-28 [*DOI :* 10.1016/J.FUTURE.2019.08.024], https://hal.inria.fr/hal-02275473

[5] B. CONFAIS, B. PARREIN, A. LEBRE.*Data Location Management Protocol for Object Stores in a Fog Computing Infrastructure*, in "IEEE Transactions on Network and Service Management", July 2019, p. 1-14 [*DOI :* 10.1109/TNSM.2019.2929823], https://hal.archives-ouvertes.fr/hal-02190125

[6] T.-D. PHAN, G. PALLEZ, S. IBRAHIM, P. RAGHAVAN.*A New Framework for Evaluating Straggler Detection Mechanisms in MapReduce*, in "ACM Transactions on Modeling and Performance Evaluation of Computing Systems", April 2019, vol. X, p. 1-22 [*DOI :* 10.1145/3328740], https://hal.inria.fr/hal-02172590

[7] S. WU, H. CHEN, H. JIN, S. IBRAHIM.*Shadow: Exploiting the Power of Choice for Efficient Shuffling in MapReduce*, in "IEEE transactions on big data", September 2019, p. 1-15 [*DOI :* 10.1109/TBDATA.2019.2943473], https://hal.inria.fr/hal-02389072

[8] A. C. ZHOU, B. SHEN, Y. XIAO, S. IBRAHIM, B. HE.*Cost-Aware Partitioning for Efficient Large Graph Processing in Geo-Distributed Datacenters*, in "IEEE Transactions on Parallel and Distributed Systems", November 2019, p. 1-1 [*DOI :* 10.1109/TPDS.2019.2955494], https://hal.inria.fr/hal-02389120

Invited Conferences

[9] M. ABDERRAHIM, M. OUZZIF, K. GUILLOUARD, J. FRANÇOIS, A. LEBRE, C. PRUD'HOMME, X. LORCA.*Efficient Resource Allocation for Multi-tenant Monitoring of Edge Infrastructures*, in "PDP 2019 - 27th Euromicro International Conference on Parallel, Distributed and Network-Based Processing", Pavie, Italy, 27th Euromicro International Conference on Parallel, Distributed and Network-Based Processing, IEEE, 2019, p. 1-8 [*DOI :* 10.1109/EMPDP.2019.8671621], https://hal.inria.fr/hal-01987946

International Conferences with Proceedings

[10] M. BELAIR, S. LANIEPCE, J.-M. MENAUD.*Leveraging Kernel Security Mechanisms to Improve Container Security: a Survey*, in "IWSECC 2019 - 2nd International Workshop on Security Engineering for Cloud Computing", Canterbury, United Kingdom, August 2019, p. 1-6 [*DOI :* 10.1145/3339252.3340502], https://hal.inria.fr/hal-02169298

[11] E. CADOREL, H. COULLON, J.-M. MENAUD.*A workflow scheduling deadline-based heuristic for energy optimization in Cloud*, in "GreenCom 2019 - 15th IEEE International Conference on Green Computing and Communications", Atlanta, United States, IEEE, July 2019, p. 1-10, https://hal.inria.fr/hal-02165835

[12] H. COULLON, C. JARD, D. LIME.*Integrated Model-checking for the Design of Safe and Efficient Distributed Software Commissioning*, in "IFM 2019 - 15th International Conference on integrated Formal Methods", Bergen, Norway, Integrated Formal Methods, December 2019, p. 120–137, https://hal.archives-ouvertes.fr/hal-02323641

[13] J. DARROUS, S. IBRAHIM, C. PÉREZ.*Is it time to revisit Erasure Coding in Data-intensive clusters?*, in "MASCOTS 2019 - 27th IEEE International Symposium on the Modeling, Analysis, and Simulation of Computer and Telecommunication Systems", Rennes, France, IEEE, October 2019, p. 165-178 [*DOI :* 10.1109/MASCOTS.2019.00026], https://hal.inria.fr/hal-02263116

[14] J. DARROUS, T. LAMBERT, S. IBRAHIM.*On the Importance of Container Image Placement for Service Provisioning in the Edge*, in "ICCCN 2019 - 28th International Conference on Computer Communications and Networks", Valencia, Spain, IEEE, July 2019, p. 1-9 [*DOI :* 10.1109/ICCCN.2019.8846920], https://hal.inria.fr/hal-02134507

[15] H. FAN, S. WU, S. IBRAHIM, X. CHEN, H. JIN, J. XIAO, H. GUAN.*NCQ-Aware I/O Scheduling for Conventional Solid State Drives*, in "IPDPS 2019 - 33rd IEEE International Parallel & Distributed Processing Symposium", Rio de Janeiro, Brazil, IEEE, May 2019, p. 523-532 [*DOI :* 10.1109/IPDPS.2019.00062], https://hal.inria.fr/hal-02389113

[16] T. L. NGUYEN, R. NOU, A. LEBRE.*YOLO: Speeding up VM and Docker Boot Time by reducing I/O operations*, in "EURO-PAR 2019 - European Conference on Parallel Processing", Göttingen, Germany, Springer, August 2019, p. 273-287 [*DOI :* 10.1007/978-3-030-29400-7_20], https://hal.inria.fr/hal-02172288

[17] H. SAHLI, T. LEDOUX, É. RUTTEN.*Modeling Self-Adaptive Fog Systems Using Bigraphs*, in "FOCLASA 2019 - 17th International Workshop on coordination and Self-Adaptativeness of Sotware applications", Oslo, Norway, September 2019, p. 1-16, https://hal.inria.fr/hal-02271394

[18] Y. WANG, D. NÖRTERSHÄUSER, S. LE MASSON, J.-M. MENAUD.*An Empirical Study of Power Characterization Approaches for Servers*, in "ENERGY 2019 - The Ninth International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies", Athens, Greece, June 2019, p. 1-6, https://hal.inria.fr/hal-02120589

[19] Y. WANG, D. NÖRTERSHÄUSER, S. LE MASSON, J.-M. MENAUD.*Experimental Characterization of Variation in Power Consumption for Processors of Different generations*, in "GreenCom 2019 - 15th IEEE International Conference on Green Computing and Communications", Atlanta, United States, IEEE, July 2019, p. 1-9, https://hal.inria.fr/hal-02166019

[20] S. WU, D. HU, S. IBRAHIM, H. JIN, J. XIAO, F. CHEN, H. LIU.*When FPGA-Accelerator Meets Stream Data Processing in the Edge*, in "ICDCS'19 - 39th IEEE International Conference on Distributed Computing

Systems", Dallas, United States, IEEE, July 2019, p. 1818-1829 [*DOI :* 10.1109/ICDCS.2019.00180], https://hal.inria.fr/hal-02389101

[21] A. C. ZHOU, Y. XIAO, B. HE, S. IBRAHIM, R. CHENG.*Incorporating Probabilistic Optimizations for Resource Provisioning of Data Processing Workflows*, in "ICPP 2019 - 48th International Conference on Parallel Processing", Kyoto, Japan, ACM Press, August 2019, p. 1-10 [*DOI :* 10.1145/3337821.3337847], https://hal.inria.fr/hal-02389078

Conferences without Proceedings

[22] F. AIT SALAHT, F. DESPREZ, A. LEBRE, C. PRUD'HOMME, M. ABDERRAHIM.*Service Placement in Fog Computing Using Constraint Programming*, in "SCC 2019 - IEEE International Conference on Services Computing", Milan, Italy, IEEE, July 2019, p. 19-27 [*DOI :* 10.1109/SCC.2019.00017], https://hal.archives-ouvertes.fr/hal-02108806

[23] M. BELAIR, S. LANIEPCE, J.-M. MENAUD.*Container interaction with host OS for enhanced security: a survey*, in "COMPAS 2019 - Conférence d'informatique en Parallélisme, Architecture et Système", Anglet, France, June 2019, https://hal.inria.fr/hal-02165816

[24] F.-Z. BOUJDAD, A. GAIGNARD, M. SÜDHOLT, W. GARZÓN-ALFONSO, L. D. BENAVIDES NAVARRO, R. REDON.*On distributed collaboration for biomedical analyses*, in "CCGrid-Life 2019 Workshop on Clusters, Clouds and Grids for Life Sciences", Larnaca, Cyprus, IEEE, May 2019, p. 1-10 [*DOI :* 10.1109/CCGRID.2019.00079], https://hal.archives-ouvertes.fr/hal-02080463

[25] D. SAINGRE, T. LEDOUX, J.-M. MENAUD.*BCTMark - Vers un outil pour l'évaluation des performances et du coût énergétique des technologies blockchain*, in "COMPAS 2019 - Conférence d'informatique en Parallélisme, Architecture et Système", Anglet, France, June 2019, https://hal.inria.fr/hal-02166044

Research Reports

[26] F. AIT SALAHT, F. DESPREZ, A. LEBRE.*An overview of service placement problem in Fog and Edge Computing*, Univ Lyon, EnsL, UCBL, CNRS, Inria, LIP, LYON, France, October 2019, n° RR-9295, p. 1-43, https://hal.inria.fr/hal-02313711

[27] A. BAUSKAR, A. DA SILVA, A. LEBRE, C. MOMMESSIN, P. NEYRON, Y. NGOKO, Y. RICORDEL, D. TRYSTRAM, A. VAN KEMPEN.*Investigating Placement Challenges in Edge Infrastructures through a Common Simulator*, DATAMOVE ; STACS ; DAPI IMT Atlantique, July 2019, n° RR-9282, p. 1-16, https://hal.inria.fr/hal-02153203

[28] T. L. NGUYEN, R. NOU, A. LEBRE.*YOLO: Speeding up VM Boot Time by reducing I/O operations*, Inria, January 2019, n° RR-9245, p. 1-18, https://hal.inria.fr/hal-01983626

Scientific Popularization

[29] J.-C. BACH, A. BEUGNARD, H. BRUNELIERE, H. COULLON, F. LEHUÉDÉ, G. MASSONNET, J. NOYÉ, G. SIMONIN.*Logiciel et Industrie du Futur*, Mathématiques et Informatique, Presse des Mines, October 2019, 62, https://hal.archives-ouvertes.fr/hal-02299214

[30] D. ESPINEL SARMIENTO, A. LEBRE, L. NUSSBAUM, A. CHARI.*Distributing connectivity management in Cloud-Edge infrastructures : Challenges and approaches*, in "COMPAS 2019 - Conférence d'informatique en Parallélisme, Architecture et Système", Anglet, France, June 2019, p. 1-7, https://hal.inria.fr/hal-02133606

Other Publications

[31] J. DARROUS, S. IBRAHIM.*Enabling Data Processing under Erasure Coding in the Fog*, August 2019, 1, ICPP 2019 - 48th International Conference on Parallel Processing, Poster, https://hal.inria.fr/hal-02388835

References in notes

[32] *Akamai Cloudlets*, 2018, (Accessed: 2018-03-08), http://cloudlets.akamai.com

[33] *Amazon Lambda@Edge*, 2018, (Accessed: 2018-03-08), https://aws.amazon.com/lambda/edge/

[34] C. ATKINSON, T. SCHULZE, S. KLINGERT.*Facilitating Greener IT through Green Specifications*, in "IEEE Software", May 2014, vol. 31, n^{o} 3, p. 56-63, http://dx.doi.org/10.1109/MS.2014.19

[35] F. BAUDE, D. CAROMEL, C. DALMASSO, M. DANELUTTO, V. GETOV, L. HENRIO, C. PÉREZ.*GCM: a grid extension to Fractal for autonomous distributed components*, in "annals of telecommunications", 2009, vol. 64, n^{o} 1-2, p. 5-24, http://dx.doi.org/10.1007/s12243-008-0068-8

[36] F. BAUDE, L. HENRIO, C. RUZ.*Programming distributed and adaptable autonomous components–the GCM/ProActive framework*, in "Software: Practice and Experience", May 2014, https://hal.inria.fr/hal-01001043

[37] 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, October 2017, https://hal.inria.fr/hal-01593144

[38] G. BLAIR, T. COUPAYE, J.-B. STEFANI.*Component-based architecture: the Fractal initiative*, in "Annals of telecommunications", February 2009, vol. 64, n^{o} 1, p. 1–4, https://doi.org/10.1007/s12243-009-0086-1

[39] P. BODIK, R. GRIFFITH, C. SUTTON, A. FOX, M. I. JORDAN, D. A. PATTERSON.*Automatic Exploration of Datacenter Performance Regimes*, in "Proceedings of the 1st Workshop on Automated Control for Datacenters and Clouds", New York, NY, USA, ACDC '09, ACM, 2009, p. 1–6, http://doi.acm.org/10.1145/1555271.1555273

[40] F. BONOMI, R. MILITO, J. ZHU, S. ADDEPALLI.*Fog computing and its role in the internet of things*, in "Proceedings of the first edition of the MCC workshop on Mobile cloud computing", ACM, 2012, p. 13–16

[41] F.-Z. BOUJDAD, M. SÜDHOLT.*Constructive Privacy for Shared Genetic Data*, in "CLOSER 2018 - 8th International Conference on Cloud Computing and Services Science", Funchal, Madeira, Portugal, Proceedings of CLOSER 2018, March 2018, p. 1-8, https://hal.archives-ouvertes.fr/hal-01692620

[42] F. BRASILEIRO, G. SILVA, F. ARAÚJO, M. NÓBREGA, I. SILVA, G. ROCHA.*Fogbow: A Middleware for the Federation of IaaS Clouds*, in "The 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)", IEEE, 2016, p. 531–534

[43] O. BRAČEVAC, S. ERDWEG, G. SALVANESCHI, M. MEZINI.*CPL: A Core Language for Cloud Computing*, in "Proceedings of the 13th International Conference on Modularity (MODULARITY 2016)", ACM Press, 2016, p. 94–105, http://doi.acm.org/10.1145/2889443.2889452

[44] H. BRUNELIERE, Z. AL-SHARA, F. ALVARES, J. LEJEUNE, T. LEDOUX.*A Model-based Architecture for Autonomic and Heterogeneous Cloud Systems*, in "CLOSER 2018 - 8h International Conference on Cloud Computing and Services Science", Funchal, Portugal, March 2018, vol. 1, p. 201-212, Best Paper Award [*DOI :* 10.5220/0006773002010212], https://hal.archives-ouvertes.fr/hal-01705248

[45] E. BRUNETON, T. COUPAYE, M. LECLERCQ, V. QUEMA, J.-B. STEFANI.*An Open Component Model and Its Support in Java*, in "Component-Based Software Engineering", Berlin, Heidelberg, I. CRNKOVIC, J. A. STAFFORD, H. W. SCHMIDT, K. WALLNAU (editors), Springer Berlin Heidelberg, 2004, p. 7–22

[46] P. CARBONE, A. KATSIFODIMOS, S. EWEN, V. MARKL, S. HARIDI, K. TZOUMAS.*Apache Flink?: Stream and Batch Processing in a Single Engine*, in "IEEE Data Eng. Bull.", 2015, vol. 38, p. 28-38

[47] V. CARDELLINI, F. L. PRESTI, M. NARDELLI, G. R. RUSSO.*Towards Hierarchical Autonomous Control for Elastic Data Stream Processing in the Fog*, in "Euro-Par 2017: Parallel Processing Workshops: Euro-Par 2017 International Workshops, Santiago de Compostela, Spain, August 28-29, 2017, Revised Selected Papers", Springer, 2018, vol. 10659, p. 106-117

[48] F.-C. CHANG, H.-C. HUANG, H.-M. HANG.*Combined Encryption and Watermarking Approaches for Scalable Multimedia Coding*, K. AIZAWA, Y. NAKAMURA, S. SATOH (editors), Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, p. 356–363

[49] R.-A. CHERRUEAU, R. DOUENCE, M. SÜDHOLT.*A Language for the Composition of Privacy-Enforcement Techniques*, in "IEEE RATSP 2015, The 2015 IEEE International Symposium on Recent Advances of Trust, Security and Privacy in Computing and Communications ", Helsinki, Finland, August 2015, https://hal.inria.fr/hal-01168303

[50] R.-A. CHERRUEAU, A. LEBRE, D. PERTIN, F. WUHIB, J. SOARES.*Edge Computing Resource Management System: a Critical Building Block! Initiating the debate via OpenStack*, in "The USENIX Workshop on Hot Topics in Edge Computing (HotEdge'18)", july 2018

[51] H. E. CHIHOUB, S. IBRAHIM, Y. LI, G. ANTONIU, M. S. PEREZ, L. BOUGE.*Exploring Energy-Consistency Trade-Offs in Cassandra Cloud Storage System*, in "27th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)", October 2015, p. 146-153

[52] B. CONFAIS, A. LEBRE, 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", 2017

[53] J. C. CORBETT, J. DEAN, M. EPSTEIN, A. FIKES, C. FROST, J. J. FURMAN, S. GHEMAWAT, A. GUBAREV, C. HEISER, P. HOCHSCHILD.*Spanner: Google's globally distributed database*, in "ACM Transactions on Computer Systems (TOCS)", 2013, vol. 31, n^{o} 3, 8

[54] R. D. COSMO, J. MAURO, S. ZACCHIROLI, G. ZAVATTARO.*Aeolus: A component model for the cloud*, in "Information and Computation", 2014, vol. 239, n^{o} Supplement C, p. 100–121 [*DOI :* 10.1016/J.IC.2014.11.002], http://www.sciencedirect.com/science/article/pii/S0890540114001424

[55] H. COULLON, J. BIGOT, C. PÉREZ.*Extensibility and Composability of a Multi-Stencil Domain Specific Framework*, in "International Journal of Parallel Programming", November 2017, https://doi.org/10.1007/s10766-017-0539-5

[56] H. COULLON, G. LE LOUËT, J.-M. MENAUD.*Virtual Machine Placement for Hybrid Cloud using Constraint Programming*, in "ICPADS 2017", 2017

[57] H. COULLON, D. PERTIN, C. PÉREZ.*Production Deployment Tools for IaaSes: an Overall Model and Survey*, in "IEEE International Conference on Future Internet of Things and Cloud (FiCloud) 2017", Prague, Czech Republic, August 2017, https://hal.inria.fr/hal-01532489

[58] 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", 2017

[59] P.-C. DAVID, T. LEDOUX, M. LÉGER, T. COUPAYE.*FPath and FScript: Language support for navigation and reliable reconfiguration of Fractal architectures*, in "annals of telecommunications - annales des télécommunications", February 2009, vol. 64, n^o 1, p. 45–63, https://doi.org/10.1007/s12243-008-0073-y

[60] F. A. DE OLIVEIRA, T. LEDOUX, R. SHARROCK.*A framework for the coordination of multiple autonomic managers in cloud environments*, in "Self-Adaptive and Self-Organizing Systems (SASO), 2013 IEEE 7th International Conference on", IEEE, 2013, p. 179–188

[61] W. FELTER, A. FERREIRA, R. RAJAMONY, J. RUBIO.*An updated performance comparison of virtual machines and linux containers*, in "Performance Analysis of Systems and Software (ISPASS), 2015 IEEE International Symposium On", IEEE, 2015, p. 171–172

[62] A. FLISSI, J. DUBUS, N. DOLET, P. MERLE.*Deploying on the Grid with DeployWare*, in "Eighth IEEE International Symposium on Cluster Computing and the Grid", France, May 2008, p. 177-184 [*DOI :* 10.1109/CCGRID.2008.59], https://hal.archives-ouvertes.fr/hal-00259836

[63] P. GARCIA LOPEZ, A. MONTRESOR, D. EPEMA, A. DATTA, T. HIGASHINO, A. IAMNITCHI, M. BARCELLOS, P. FELBER, E. RIVIERE.*Edge-centric Computing: Vision and Challenges*, in "SIGCOMM Comput. Commun. Rev.", September 2015, vol. 45, n^o 5, p. 37–42, http://doi.acm.org/10.1145/2831347.2831354

[64] Í. GOIRI, W. KATSAK, K. LE, T. D. NGUYEN, R. BIANCHINI.*Parasol and GreenSwitch: Managing Datacenters Powered by Renewable Energy*, in "SIGARCH Comput. Archit. News", March 2013, vol. 41, n^o 1, p. 51–64

[65] Í. GOIRI, K. LE, T. D. NGUYEN, J. GUITART, J. TORRES, R. BIANCHINI.*GreenHadoop: Leveraging Green Energy in Data-processing Frameworks*, in "Proceedings of the 7th ACM European Conference on Computer Systems", New York, NY, USA, EuroSys '12, ACM, 2012, p. 57–70, http://doi.acm.org/10.1145/2168836.2168843

[66] H. GOUDARZI, M. PEDRAM.*Geographical Load Balancing for Online Service Applications in Distributed Datacenters*, in "in IEEE international conference on cloud computing (CLOUD 2013", 2013

[67] L. GU, D. ZENG, A. BARNAWI, S. GUO, I. STOJMENOVIC.*Optimal Task Placement with QoS Constraints in Geo-Distributed Data Centers Using DVFS*, in "IEEE Transactions on Computers", July 2015, vol. 64, n^{o} 7, p. 2049-2059, http://dx.doi.org/10.1109/TC.2014.2349510

[68] K. HA, Y. ABE, T. EISZLER, Z. CHEN, W. HU, B. AMOS, R. UPADHYAYA, P. PILLAI, M. SATYANARAYANAN.*You Can Teach Elephants to Dance: Agile VM Handoff for Edge Computing*, in "Proceedings of the Second ACM/IEEE Symposium on Edge Computing", New York, NY, USA, SEC '17, ACM, 2017, http://doi.acm.org/10.1145/3132211.3134453

[69] B. HAN, V. GOPALAKRISHNAN, L. JI, S. LEE.*Network function virtualization: Challenges and opportunities for innovations*, in "IEEE Communications Magazine", February 2015, vol. 53, n^{o} 2, p. 90-97, http://dx.doi.org/10.1109/MCOM.2015.7045396

[70] 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", April 2017, vol. 2, n^{o} 2, p. 113-126, http://dx.doi.org/10.1109/TSUSC.2017.2714959

[71] F. HERMENIER, J. LAWALL, G. MULLER.*Btrplace: A flexible consolidation manager for highly available applications*, in "IEEE Transactions on dependable and Secure Computing", 2013, vol. 10, n^{o} 5, p. 273–286

[72] F. HERMENIER, A. LEBRE, J.-M. MENAUD.*Cluster-wide Context Switch of Virtualized Jobs*, in "Proceedings of the Virtualization Technologies in Distributed Computing Workshop (co-locaed with ACM HPDC'10)", New York, NY, USA, HPDC '10, ACM, 2010, p. 658–666, http://doi.acm.org/10.1145/1851476.1851574

[73] F. HERMENIER, X. LORCA, J.-M. MENAUD, G. MULLER, J. LAWALL.*Entropy: a consolidation manager for clusters*, in "Proceedings of the 2009 ACM SIGPLAN/SIGOPS international conference on Virtual execution environments", ACM, 2009, p. 41–50

[74] M. HIRZEL, S. SCHNEIDER, B. GEDIK.*SPL: An Extensible Language for Distributed Stream Processing*, in "toplas", March 2017, vol. 39, n^{o} 1, http://doi.acm.org/10.1145/3039207

[75] S. IBRAHIM, B. HE, H. JIN.*Towards Pay-As-You-Consume Cloud Computing*, in "2011 IEEE International Conference on Services Computing", July 2011, p. 370-377, http://dx.doi.org/10.1109/SCC.2011.38

[76] S. IFTIKHAR, S. KHAN, Z. ANWAR, OTHERS.*GenInfoGuard: A Robust and Distortion-Free Watermarking Technique for Genetic Data*, in "PLOS ONE", February 2015, vol. 10, n^{o} 2, p. 1-22, https://doi.org/10.1371/journal.pone.0117717

[77] B. JENNINGS, R. STADLER.*Resource Management in Clouds: Survey and Research Challenges*, in "Journal of Network and Systems Management", July 2015, vol. 23, n^{o} 3, p. 567–619, https://doi.org/10.1007/s10922-014-9307-7

[78] E. JONARDI, M. A. OXLEY, S. PASRICHA, A. A. MACIEJEWSKI, H. J. SIEGEL.*Energy cost optimization for geographically distributed heterogeneous data centers*, in "2015 Sixth International Green and Sustainable Computing Conference (IGSC)", December 2015, p. 1-6, http://dx.doi.org/10.1109/IGCC.2015.7393677

[79] H. KARAU, A. KONWINSKI, P. WENDELL, M. ZAHARIA.*Learning Spark*, O'Reilly Media, February 2015

[80] J. KRAMER, J. MAGEE.*The evolving philosophers problem: dynamic change management*, in "IEEE Transactions on Software Engineering", November 1990, vol. 16, n^{o} 11, p. 1293-1306, http://dx.doi.org/10.1109/32.60317

[81] A. LEBRE, J. PASTOR, A. SIMONET, F. DESPREZ.*Revising OpenStack to Operate Fog/Edge Computing Infrastructures*, in "The IEEE International Conference on Cloud Engineering (IC2E)", April 2017, p. 138-148, http://dx.doi.org/10.1109/IC2E.2017.35

[82] G. MALEWICZ, M. H. AUSTERN, A. J. BIK, J. C. DEHNERT, I. HORN, N. LEISER, G. CZAJKOWSKI.*Pregel: A System for Large-scale Graph Processing*, in "Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data", New York, NY, USA, SIGMOD '10, ACM, 2010, p. 135–146, http://doi.acm.org/10.1145/1807167.1807184

[83] T. L. NGUYEN, A. LEBRE.*Virtual Machine Boot Time Model*, in "Parallel, Distributed and Network-based Processing (PDP), 2017 25th Euromicro International Conference on", IEEE, 2017, p. 430–437

[84] K. OKAMURA, Y. OYAMA.*Load-based covert channels between Xen virtual machines*, in "Proceedings of the 2010 ACM Symposium on Applied Computing", ACM, 2010, p. 173–180

[85] P. OREIZY, N. MEDVIDOVIC, R. N. TAYLOR.*Runtime Software Adaptation: Framework, Approaches, and Styles*, in "Companion of the 30th International Conference on Software Engineering", New York, NY, USA, ICSE Companion '08, ACM, 2008, p. 899–910, http://doi.acm.org/10.1145/1370175.1370181

[86] T. D. PHAN, S. IBRAHIM, G. ANTONIU, L. BOUGE.*On Understanding the Energy Impact of Speculative Execution in Hadoop*, in "2015 IEEE International Conference on Data Science and Data Intensive Systems", December 2015, p. 396-403

[87] F. QUESNEL, A. LEBRE, M. SÜDHOLT.*Cooperative and reactive scheduling in large-scale virtualized platforms with DVMS*, in "Concurrency and Computation: Practice and Experience", 2013, vol. 25, n^{o} 12, p. 1643–1655

[88] D. SABELLA, A. VAILLANT, P. KUURE, U. RAUSCHENBACH, F. GIUST.*Mobile-Edge Computing Architecture: The role of MEC in the Internet of Things*, in "IEEE Consumer Electronics Magazine", October 2016, vol. 5, n^{o} 4, p. 84-91, http://dx.doi.org/10.1109/MCE.2016.2590118

[89] D. SERRANO, S. BOUCHENAK, Y. KOUKI, F. A. DE OLIVEIRA JR., T. LEDOUX, J. LEJEUNE, J. SOPENA, L. ARANTES, P. SENS.*SLA guarantees for cloud services*, in "Future Generation Computer Systems", 2016, vol. 54, n^{o} Supplement C, p. 233–246 [*DOI :* 10.1016/J.FUTURE.2015.03.018], http://www.sciencedirect.com/science/article/pii/S0167739X15000801

[90] Q. SHEN, X. LIANG, X. S. SHEN, X. LIN, H. Y. LUO.*Exploiting Geo-Distributed Clouds for a E-Health Monitoring System With Minimum Service Delay and Privacy Preservation*, in "IEEE Journal of Biomedical and Health Informatics", March 2014, vol. 18, n^{o} 2, p. 430-439, http://dx.doi.org/10.1109/JBHI.2013.2292829

[91] W. SHI, T. F. WENISCH.*Energy-Efficient Data Centers*, in "IEEE Internet Computing", 2017, vol. 21, n^{o} 4, p. 6-7, http://dx.doi.org/10.1109/MIC.2017.2911429

[92] J. STEFANI.*Components as Location Graphs*, in "Formal Aspects of Component Software - 11th International Symposium, FACS 2014, Bertinoro, Italy, September 10-12, 2014, Revised Selected Papers", 2014, p. 3–23, https://doi.org/10.1007/978-3-319-15317-9_1

[93] C. SZYPERSKI.*Component Software: Beyond Object-Oriented Programming*, 2nd, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2002

[94] Y. TALEB, S. IBRAHIM, G. ANTONIU, T. CORTES.*Characterizing Performance and Energy-Efficiency of the RAMCloud Storage System*, in "2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)", June 2017, p. 1488-1498, http://dx.doi.org/10.1109/ICDCS.2017.51

[95] THE LINUX FOUNDATION.*Open Network Automation Platform*, 2018, (Accessed: 2018-03-08), https://www.onap.org/

[96] THE OPENSTACK FOUNDATION.*Cloud Edge Computing: Beyond the Data Center (White Paper)*, January 2018, (Accessed: 2018-03-08), https://www.openstack.org/assets/edge/OpenStack-EdgeWhitepaper-v3-online.pdf

[97] A. THUSOO, Z. SHAO, S. ANTHONY, D. BORTHAKUR, N. JAIN, J. SEN SARMA, R. MURTHY, H. LIU.*Data Warehousing and Analytics Infrastructure at Facebook*, in "Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data", SIGMOD '10, ACM Press, 2010, p. 1013–1020, http://doi.acm.org/10.1145/1807167.1807278

[98] T. WHITE.*Hadoop: The Definitive Guide*, 4th, O'Reilly Media, April 2015

[99] A. WIERMAN, Z. LIU, I. LIU, H. MOHSENIAN-RAD.*Opportunities and challenges for data center demand response*, in "International Green Computing Conference", November 2014, p. 1-10, http://dx.doi.org/10.1109/IGCC.2014.7039172

[100] J. XIAO, Z. XU, H. HUANG, H. WANG.*Security implications of memory deduplication in a virtualized environment*, in "2013 43rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)", IEEE, 2013, p. 1–12

[101] F. XU, F. LIU, H. JIN, A. V. VASILAKOS.*Managing performance overhead of virtual machines in cloud computing: A survey, state of the art, and future directions*, in "Proceedings of the IEEE", 2014, vol. 102, n^o 1, p. 11–31

[102] O. YILDIZ, M. DORIER, S. IBRAHIM, R. ROSS, G. ANTONIU.*On the Root Causes of Cross-Application I/O Interference in HPC Storage Systems*, in "2016 IEEE International Parallel and Distributed Processing Symposium (IPDPS)", May 2016, p. 750-759, http://dx.doi.org/10.1109/IPDPS.2016.50

[103] O. YILDIZ, A. C. ZHOU, S. IBRAHIM.*Eley: On the Effectiveness of Burst Buffers for Big Data Processing in HPC Systems*, in "2017 IEEE International Conference on Cluster Computing (CLUSTER)", September 2017, p. 87-91, http://dx.doi.org/10.1109/CLUSTER.2017.73

[104] M. ZAHARIA, T. DAS, H. LI, T. HUNTER, S. SHENKER, I. STOICA.*Discretized Streams: Fault-tolerant Streaming Computation at Scale*, in "Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles", New York, NY, USA, SOSP '13, ACM, 2013, p. 423–438, http://doi.acm.org/10.1145/2517349.2522737

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

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
A1.5. - Complex systems
A2.3. - Embedded and cyber-physical systems
A2.4. - Formal method for verification, reliability, certification
A2.4.2. - Model-checking
A4.5. - Formal methods for security
A6.4.3. - Observability and Controlability
A6.4.6. - Optimal control
A7.1.1. - Distributed algorithms
A7.2. - Logic in Computer Science
A8.2. - Optimization
A8.6. - Information theory
A8.11. - Game Theory

Other Research Topics and Application Domains:

B5.2.2. - Railway
B6.2. - Network technologies
B6.3.3. - Network Management
B7.1. - Traffic management
B8.5.2. - Crowd sourcing

1. Team, Visitors, External Collaborators

Research Scientists

Éric Badouel [Inria, Researcher, HDR]
Nathalie Bertrand [Team leader from Apr 2019, Inria, Researcher, HDR]
Éric Fabre [Team leader until Mar 2019, Inria, Senior Researcher, HDR]
Blaise Genest [CNRS, Senior 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]

Faculty Member

Hugo Bazille [Univ Rennes 1, ATER, from Sept 2019]

External Collaborator

Reiya Noguchi [Mitsubishi Electric]

PhD Students

Hugo Bazille [Inria, PhD Student, until Aug 2019]

Sihem Cherrared [Orange Labs, PhD Student, granted by CIFRE]
Emily Clément [Mitsubishi Electric, PhD Student, granted by CIFRE]
Arij Elmajed [Nokia, PhD Student, granted by CIFRE]
Léo Henry [Univ de Rennes I, PhD Student]
Anirban Majumdar [CNRS, PhD Student]
Abdul Majith Noordheen [Inria, PhD Student]
Victor Roussanaly [Univ de Rennes I, PhD Student]
Suman Sadhukhan [Inria, PhD Student]
Bastien Thomas [Univ de Rennes I, PhD Student, from Oct 2019]

Post-Doctoral Fellow

Adrian Puerto Aubel [Inria, Post-Doctoral Fellow, from Mar 2019]

Visiting Scientists

Khushraj Nanik Madnani [IIT Bombay, from May 2019 until Jun 2019]
Shauna Laurene Ricker [Mount Allison University, from Apr 2019 until Jul 2019]

Administrative Assistant

Laurence Dinh [Inria, Administrative Assistant]

2. Overall Objectives

2.1. Context

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.

2.2. 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.

2.3. 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.

2.4. 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. Introduction

Since its creation in 2015, SUMO has successfully developed formal methods for large quantitative systems, in particular addressing verification, synthesis and control problems. Our current motivation is to expand this by putting emphasis on new concerns, such as algorithm efficiency, imprecision handling, and the more challenging objective of addressing incomplete or missing models. In the following we list a selection of detailed research goals, structured into four axes according to model classes: quantitative models, large systems, population models, and data-driven models. Some correspond to the pursuit of previously obtained results, others are more prospective.

3.2. Axis 1: Quantitative models

The analysis and control of quantitative models will remain at the heart of a large part of our research activities. In particular, we have two starting collaborative projects focusing on **timed models**, namely our ANR project TickTac and our collaboration with MERCE. The main expected outcome of TickTac is an open-source tool implementing the latest algorithms and allowing for quick prototyping of new algorithms. Several other topics will be explored in these collaborations, including robustness issues, game-theoretic problems, as well as the development of efficient algorithms, *e.g.* based on CEGAR approach or specifically designed for subclasses of automata (*e.g.* automata with few clocks and/or having a specific structure, as in [38]). Inspired by our collaboration with Alstom, we also aim at developing symbolic techniques for analysing non-linear timed models.

Stochastic models are another important focus for our research. On the one hand, we want to pursue our work on the optimization of non-standard properties for Markov decision processes, beyond the traditional verification questions, and explore *e.g.* long-run probabilities, and quantiles. Also, we aim at lifting our work on decisiveness from purely stochastic [36], [37] to non-deterministic and stochastic models in order to provide approximation schemes for the probability of (repeated) reachability properties in infinite-state Markov decision processes. On the other hand, in order to effectively handle large stochastic systems, we will pursue our work on approximation techniques. We aim at deriving simpler models, enjoying or preserving specific properties, and at determining the appropriate level of abstraction for a given system. One needs of course to quantify the approximation degrees (distances), and to preserve essential features of the original systems (explainability). This is a connection point between formal methods and the booming learning methods.

Regarding **diagnosis/opacity** issues, we will explore further the quantitative aspects. For diagnosis, the theory needs extensions to the case of incomplete or erroneous models, and to reconfigurable systems, in order to develop its applicability (see Sec. 3.6). There is also a need for non-binary causality analysis (*e.g.* performance degradations in complex systems). For opacity, we aim at quantifying the effort attackers must produce *vs* how much of a secret they can guess. We also plan to synthesize robust controllers resisting to sensor failures/attacks.

3.3. Axis 2: Large systems

Part of the background of SUMO is on the analysis and management of concurrent and modular/distributed systems, that we view as two main approaches to address state explosion problems. We will pursue the study of these models (including their quantitative features): verification of timed concurrent systems, robust distributed control of modular systems, resilient control to coalitions of attackers, distributed diagnosis, modular opacity analysis, distributed optimal planning, etc. Nevertheless, we have identified two new lines of effort, inspired by our application domains.

Reconfigurable systems. This is mostly motivated by applications at the convergence of virtualization techs with networking (Orange and Nokia PhDs). Software defined networks, either in the core (SDN/NFV) or at the edge (IoT) involve distributed systems that change structure constantly, to adapt to traffic, failures, maintenance, upgrades, etc. Traditional verification, control, diagnosis approaches (to mention only those) assume static and known models that can be handled as a whole. This is clearly insufficient here: one needs to adapt existing results to models that (sometimes automatically) change structure, incorporate new components/users or lose some, etc. At the same time, the programming paradigms for such systems (chaos monkey) incorporate resilience mechanisms, that should be considered by our models.

Hierarchical systems. Our experience with the regulation of subway lines (Alstom) revealed that large scale complex systems are usually described at a single level of granularity. Determining the appropriate granularity is a problem in itself. The control of such systems, with humans in the loop, can not be expressed at this single level, as tasks become too complex and require extremely skilled staff. It is rather desirable to describe models simultaneously at different levels of granularity, and to perform control at the appropriate level: humans in charge of managing the system by high level objectives, and computers in charge of implementing the appropriate micro-control sequences to achieve these tasks.

3.4. Axis 3: Population models

We want to step up our effort in parameterized verification of systems consisting of many identical components, so-called population models. In a nutshell our objectives summarize as "from Boolean to quantitative".

Inspired by our experience on the analysis of populations of yeasts, we aim at developping the quantitative analysis and control of population models, *e.g.* using Markov decision processes together with quantitative properties, and focusing on generating strategies with fast convergence.

As for broadcast networks, the challenge is to model the mobility of nodes (representing mobile ad hoc networks) in a faithful way. The obtained model should reflect on the one hand, the placement of nodes at a given time instant, and on the other hand, the physical movement of nodes over time. In this context, we will also use game theory techniques which allows one to study cooperative and conflictual behaviors of the nodes in the network, and to synthesize correct-by-design systems in adversarial environments.

As a new application area, we target randomized distributed algorithms. Our goal is to provide probabilistic variants of threshold automata [39] to represent fault-tolerant randomized distributed algorithms, designed for instance to solve the consensus problem. Most importantly, we then aim at developing new parameterized verification techniques, that will enable the automated verification of the correctness of such algorithms, as well as the assessment of their performances (in particular the expected time to termination).

In this axis, we will investigate whether fluid model checking and mean-field approximation techniques apply to our problems. More generally, we aim at a fruitful cross-fertilizing of these approaches with parameterized model-checking algorithms.

3.5. Axis 4: Data-driven models

In this axis, we will consider data-centric models, and in particular their application to crowd-sourcing. Many data-centric models such as Business Artifacts [40] orchestrate simple calls and answers to tasks performed by a single user. In a crowd-sourcing context, tasks are realized by pools of users, which may result in imprecise, uncertain and (partially) incompatible information. We thus need mechanisms to reconcile and fuse the various contributions in order to produce reliable information. Another aspect to consider concerns answers of higher-order: how to allow users to return intentional answers, under the form of a sub-workflow (coordinated set of tasks) which execution will provide the intended value. In the framework of the ANR Headwork we will build on formalisms such as GAG (guarded attribute grammars) or variants of business artifacts to propose formalisms adapted to crowd-sourcing applications, and tools to analyze them. To address imprecision, we will study techniques to handle fuzziness in user answers, will explore means to set incentives (rewards) dynamically, and to set competence requirements to guide the execution of a complex workflow, in order to achieve an objective with a desired level of quality.

In collaboration with Open Agora, CESPA and University of Yaoundé (Cameroun) we intend to implement in the GAG formalism some elements of argumentation theory (argumentation schemes, speech acts and dialogic games) in order to build a tool for the conduct of a critical discussion and the collaborative construction of expertise. The tool would incorporate point of view extraction (using clustering mechanisms), amendment management and consensus building mechanisms.

3.6. Transversal concern: missing models

We are concerned with one important lesson derived from our involvement in several application domains. Most of our background gets in force as soon as a perfect model of the system under study is available. Then verification, control, diagnosis, test, etc. can mobilize a solid background, or suggest new algorithmic problems to address. In numerous situations, however, assuming that a model is available is simply unrealistic. This is a major bottleneck for the impact of our research. We therefore intend to address this difficulty, in particular for the following domains.

- Model building for diagnosis. As a matter of fact, diagnosis theory hardly touches the ground to the extent that complete models of normal behavior are rarely available, and the identification of the appropriate abstraction level is unclear. Knowledge of faults and their effects is even less accessible. Also, the actual implemented systems may differ significantly from behaviors described in the norms. One therefore needs a theory for incomplete and erroneous models. Besides, one is often less bothered by partial observations than drowned by avalanches of alerts when malfunctions occur. Learning may come to the rescue, all the more that software systems may be deployed in sandpits and damaged for experimentation, thus allowing the collection of masses of labeled data. Competition on that theme clearly comes from Machine Learning techniques.

- Verification of large scale software. For some verification problems like the one we address in the IPL HAC-Specis, one does not have access to a formal model of the distributed program under study, but only to executions in a simulator. Formal verification poses new problems due to the difficulties to capture global states, to master state space explosion by gathering and exploiting concurrency information.
- Learning of stochastic models. Applications in bioinformatics often lead to large scale models, involving numerous chains of interactions between chemical species and/or cells. Fine grain models can be very precise, but very inefficient for inference or verification. Defining the appropriate levels of description/abstraction, given the available data and the verification goals, remains an open problem. This cannot be considered as a simple data fitting problem, as elements of biological knowledge must be combined with the data in order to preserve explainability of the phenomena.
- Testing and learning timed models: during conformance testing of a black-box implementation against its formal specification, one wants to detect non-conformances but may also want to learn the implementation model. Even though mixing testing and learning is not new, this is more recent and challenging for continuous-time models.
- Process mining. We intend to extend our work on process discovery using Petri net synthesis [35] by using negative information (*e.g.* execution traces identified as outliers) and quantitative information (probabilistic or fuzzy sets of execution traces) in order to infer more robust and precise models.

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 regulation 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.

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 crowd sourcing applications. All these applications are data-centric and user-driven. They are often 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 crowd sourcing applications 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 build prototypes to experiment them on real use cases from contributive science, health-management systems, and crowd sourcing applications. We also plan to define abstaction schemes allowing formal reasonning on these systems.

5. Highlights of the Year

5.1. Changes in 2019

SUMO was evaluated in spring 2019, and we took this opportunity to make several changes. First, we adapted the research axes of the team in our scientific foundations to reflect a slight topic drift over the last four years, which is also a consequence of modifications in the team composition. In particular, we now put emphasis on one emergent topic, namely population models. Last but not least, Éric Fabre stepped down as project-team leader and Nathalie Bertrand replaces him since April 2019.

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. 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 objectives 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. From 2015 to 2018, it was used for the P22 collaboration with Asltom Transport, to model metro traffic and evaluate performance of regulation solutions. It is now (2019) at the heart of a collaboration on multi-modal networks with Alstom transport Madrid. 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.

Since then, 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. In 2019, it was extended to include planning-based regulation techniques during a collaboration with Roma 3 University.

RELEASE FUNCTIONAL DESCRIPTION: modeling of continuous vehicles movements

- Participants: Abd El Karim Kecir and Loïc Hélouët
- Contact: Loïc Hélouët
- URL: http://www.irisa.fr/sumo/Software/SIMSTORS/

7. New Results

7.1. New results on Axis 1: Quantitative models

7.1.1. Verification of Real-Time Models

Participants : Ocan Sankur, Nicolas Markey, Victor Roussanaly

7.1.1.1. Abstraction-refinement algorithms for model checking of timed automata.

The abstraction domain we consider [26] abstracts away zones by restricting the set of clock constraints that can be used to define them, while the refinement procedure computes the set of constraints that must be taken into consideration in the abstraction so as to exclude a given spurious counterexample. We implement this idea in two ways: an enumerative algorithm where a lazy abstraction approach is adopted, meaning that possibly different abstract domains are assigned to each exploration node; and a symbolic algorithm where the abstract transition system is encoded with Boolean formulas.

7.1.1.2. Robust controller synthesis problem in Büchi timed automata

We solve a robust controller synthesis problem [20] in a purely symbolic way. The goal of the controller is to play according to an accepting lasso of the automaton, while resisting to timing perturbations chosen by a competing environment. The problem was previously shown to be *PSPACE*-complete using regions-based techniques, but we provide a first tool solving the problem using zones only, thus more resilient to state-space explosion problem. The key ingredient is the introduction of branching constraint graphs allowing to decide in polynomial time whether a given lasso is robust, and even compute the largest admissible perturbation if it is. We also make an original use of constraint graphs in this context in order to test the inclusion of timed reachability relations, crucial for the termination criterion of our algorithm. Our techniques are illustrated using a case study on the regulation of a train network.

7.1.2. Verification of Stochastic Models

Participants : Hugo Bazille, Nathalie Bertrand, Éric Fabre, Blaise Genest, Ocan Sankur

7.1.2.1. Long-run satisfaction of path properties

We introduced the concepts of long-run frequency of path properties for paths in Kripke structures, and their generalization to long-run probabilities for schedulers in Markov decision processes [13]. We then studied the natural optimization problem of computing the optimal values of these measures, when ranging over all paths or all schedulers, and the corresponding decision problem when given a threshold. The main results are as follows. For (repeated) reachability and other simple properties, optimal long-run probabilities and corresponding optimal memoryless schedulers are computable in polynomial time. When it comes to constrained reachability properties, memoryless schedulers are no longer sufficient, even in the non-probabilistic setting. Nevertheless, optimal long-run probabilities for constrained reachability are computable in pseudo-polynomial time in the probabilistic setting and in polynomial time for Kripke structures. Finally for co-safety properties expressed by NFA, we gave an exponential-time algorithm to compute the optimal long-run frequency, and proved the PSPACE-completeness of the threshold problem.

7.1.2.2. Approximate Verification of Dynamic Bayesian Networks.

We are interested in studying the evolution of large homogeneous populations of cells, where each cell is assumed to be composed of a group of biological players (species) whose dynamics is governed by a complex biological pathway, identical for all cells. Modeling the inherent variability of the species concentrations in different cells is crucial to understand the dynamics of the population. In [9], we focus on handling this variability by modeling each species by a random variable that evolves over time. This appealing approach runs into the curse of dimensionality since exactly representing a joint probability distribution involving a large set of random variables quickly becomes intractable as the number of variables grows. To make this approach amenable to biopathways, we explore different techniques to (i) approximate the exact joint distribution at a given time point, and (ii) to track its evolution as time elapses.

7.1.2.3. Classification among stochastic systems

An important task in AI is one of classifying an observation as belonging to one class among several (e.g. image classification). We revisit this problem in a verification context: given k partially observable systems modeled as Hidden Markov Models (HMMs, also called labeled Markov chains), and an execution of one of them, can we eventually classify which system performed this execution, just by looking at its observations? Interestingly, this problem generalizes several problems in verification and control, such as fault diagnosis and opacity. Also, classification has strong connections with different notions of distances between stochastic models.

In [12], we study a general and practical notion of classifiers, namely limit-sure classifiers, which allow misclassification, i.e. errors in classification, as long as the probability of misclassification tends to 0 as the length of the observation grows. To study the complexity of several notions of classification, we develop techniques based on a simple but powerful notion of stationary distributions for HMMs. We prove that one cannot classify among HMMs iff there is a finite separating word from their stationary distributions. This provides a direct proof that classifiability can be checked in PTIME, as an alternative to existing proofs using separating events (i.e. sets of infinite separating words) for the total variation distance. Our approach also allows us to introduce and tackle new notions of classifiability which are applicable in a security context.

7.1.2.4. Fault diagnosis for stochastic systems

Diagnosis of partially observable stochastic systems prone to faults was introduced in the late nineties. Diagnosability, *i.e.* the existence of a diagnoser, may be specified in different ways: exact diagnosability requires that almost surely a fault is detected and that no fault is erroneously claimed; approximate diagnosability tolerates a small error probability when claiming a fault; last, accurate approximate diagnosability guarantees that the error probability can be chosen arbitrarily small.

In the article [7], we first refine the specification of diagnosability by identifying three criteria: (1) detecting faulty runs or providing information for all runs (2) considering finite or infinite runs, and (3) requiring or not a uniform detection delay. We then give a complete picture of relations between the different diagnosability specifications for probabilistic systems and establish characterisations for most of them in the finite-state case. Based on these characterisations, we develop decision procedures, study their complexity and prove their optimality. We also design synthesis algorithms to construct diagnosers and we analyse their memory requirements. Finally we establish undecidability of the diagnosability problems for which we provided no characterisation.

7.1.3. Energy Games

Participants : Loïc Hélouët, Nicolas Markey

7.1.3.1. Games with reachability objectives under energy constraints.

Under strict energy constraints (either only lower-bound constraint or interval constraint), we prove [23] that games with reachability objectives are LOGSPACE-equivalent to energy games with the same energy constraints but without reachability objective (i.e., for infinite runs). We then consider two kinds of relaxations of the upper-bound constraints (while keeping the lower-bound constraint strict): in the first one, called weak upper bound, the upper bound is absorbing, in the sense that it allows receiving more energy when the upper bound is already reached, but the extra energy will not be stored; in the second one, we allow for temporary violations of the upper bound, imposing limits on the number or on the amount of violations. We prove that when considering weak upper bound, reachability objectives require memory, but can still be solved in polynomial-time for one-player arenas; we prove that they are in co-NP in the two-player setting. Allowing for bounded violations makes the problem PSPACE-complete for one-player arenas and EXPTIME-complete for two players.

7.2. New results on Axis 2: Large Systems Models

7.2.1. Smart Transportation Systems

Participants : Nathalie Bertrand, Loïc Hélouët, Ocan Sankur

7.2.1.1. Smart regulation of urban train systems.

We have considered application of model checking techniques to evaluate performances of urban train systems [15]. Metros are subject to unexpected delays due to weather conditions, incidents, passenger misconduct, etc. To recover from delays and avoid their propagation to the whole network, metro operators use regulation algorithms that adapt speeds and departure dates of trains. Regulation algorithms are ad-hoc tools tuned to cope with characteristics of tracks, rolling stock, and passengers habits. However, there is no universal optimal regulation adapted in any environment. So, performance of a regulation must be evaluated before its integration in a network. In this work, we use probabilistic model-checking to evaluate the performance of regulation algorithms in simple metro lines. We model the moves of trains and random delays with Markov decision processes, and regulation as a controller that forces a decision depending on its partial knowledge of the state of the system. We then use the probabilistic model checker PRISM to evaluate performance of regulation: We compute the probability to reach a stable situation from an unstable one in less than d time units, letting d vary in a large enough time interval. This approach is applied on a case study, the metro network of Glasgow.

7.2.2. Supervisory Control

Participants : Hervé Marchand

7.2.2.1. Towards resilient supervisors against sensor deception attacks.

As a security problem, we considered in [24] feedback control systems where sensor readings may be compromised by a malicious attacker intent on causing damage to the system. We study this problem at the supervisory layer of the control system, using discrete event systems techniques. We assume that the attacker can edit the outputs from the sensors of the system before they reach the supervisory controller. In this context, we formulate the problem of synthesizing a supervisor that is robust against a large class of edit attacks on the sensor readings. The solution methodology is based on the solution of a partially observed supervisory control problem with arbitrary control patterns.

7.2.3. Multi-agent systems

Participants : Arthur Queffelec, Nicolas Markey, Ocan Sankur

7.2.3.1. Multi-agent path planning problems.

We are motivated by the increasing appeal of robots in information-gathering missions. In the problems we study [21], [22], the agents must remain interconnected. We model an area by a topological graph specifying the movement and the connectivity constraints of the agents. We study the theoretical complexity of the reachability and the coverage problems of a fleet of connected agents on various classes of topological graphs. We establish the complexity of these problems on known classes, and introduce a new class called sight-moveable graphs which admit efficient algorithms.

7.2.3.2. Quantitative semantics for Strategy Logic

We introduce and study SL[F], a quantitative extension of SL (Strategy Logic) [19], one of the most natural and expressive logics describing strategic behaviours. The satisfaction value of an SL[F] formula is a real value in [0,1], reflecting "how much" or "how well" the strategic on-going objectives of the underlying agents are satisfied. We demonstrate the applications of SL[F] in quantitative reasoning about multi-agent systems, by showing how it can express concepts of stability in multi-agent systems, and how it generalises some fuzzy temporal logics. We also provide a model-checking algorithm for our logic, based on a quantitative extension of Quantified CTL.

7.3. New results on Axis 3: Population Models

7.3.1. Verification

Participants : Nathalie Bertrand, Anirban Majumdar

7.3.1.1. Networks of many identical agents communicating by broadcast.

Broadcast networks allow one to model networks of identical nodes communicating through message broadcasts [17]. Their parameterized verification aims at proving a property holds for any number of nodes, under any communication topology, and on all possible executions. We focus on the coverability problem which dually asks whether there exists an execution that visits a configuration exhibiting some given state of the broadcast protocol. Coverability is known to be undecidable for static networks, i.e. when the number of nodes and communication topology is fixed along executions. In contrast, it is decidable in PTIME when the communication topology may change arbitrarily along executions, that is for reconfigurable networks. Surprisingly, no lower nor upper bounds on the minimal number of nodes, or the minimal length of covering execution in reconfigurable networks, appear in the literature. We showed tight bounds for cutoff and length, which happen to be linear and quadratic, respectively, in the number of states of the protocol. We also introduced an intermediary model with static communication topology and non-deterministic message losses upon sending. We showed that the same tight bounds apply to lossy networks, although, reconfigurable executions may be linearly more succinct than lossy executions. Finally, we showed NP-completeness for the natural optimisation problem associated with the cutoff.

7.3.1.2. Randomized distributed algorithms for consensus.

Randomized fault-tolerant distributed algorithms pose a number of challenges for automated verification: (i) parameterization in the number of processes and faults, (ii) randomized choices and probabilistic properties, and (iii) an unbounded number of asynchronous rounds. This combination makes verification hard. Challenge (i) was recently addressed in the framework of threshold automata. We extended threshold automata to model randomized consensus algorithms that perform an unbounded number of asynchronous rounds. For non-probabilistic properties, we showed [18] that it is necessary and sufficient to verify these properties under round-rigid schedules, that is, schedules where processes enter round r only after all processes finished round $r-1$. For almost-sure termination, we analyzed these algorithms under round-rigid adversaries, that is, fair adversaries that only generate round-rigid schedules. This allowed us to do compositional and inductive reasoning that reduces verification of the asynchronous multi-round algorithms to model checking of a one-round threshold automaton. We applied this framework and automatically verified the following classic algorithms: Ben-Or's and Bracha's seminal consensus algorithms for crashes and Byzantine faults, 2-set agreement for crash faults, and RS-Bosco for the Byzantine case.

7.3.2. Control

Participants : Nathalie Bertrand, Blaise Genest, Anirban Majumdar

7.3.2.1. Controlling a population

We introduced a new setting where a population of agents [6], 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 studied 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 proved that the population control problem is decidable, and it is a EXPTIME-complete problem. As far as we know, this is one of the first results on the control of parameterized systems. Our algorithm, which is 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. The winning strategies produced by our algorithm are optimal with respect to the synchronisation time: the maximal number of steps before synchronisation of all agents in the target state is at most polynomial in the number of agents m, and exponential in the size of the NFA. We also showed 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.

7.3.2.2. Concurrent multiplayer games with arbitrary many players

Traditional concurrent games on graphs involve a fixed number of players, who take decisions simultaneously, determining the next state of the game. In [16], we introduced a parameterized variant of concurrent games on graphs, where the parameter is precisely the number of players. Parameterized concurrent games are described by finite graphs, in which the transitions bear regular languages to describe the possible move combinations that lead from one vertex to another. We considered the problem of determining whether the first player, say Eve, has a strategy to ensure a reachability objective against any strategy profile of her opponents as a coalition. In particular Eve's strategy should be independent of the number of opponents she actually has. Technically, we focused on an *a priori* simpler setting where the languages labeling transitions only constrain the number of opponents (but not their precise action choices). These constraints are described as semilinear sets, finite unions of intervals, or intervals. We established the precise complexities of the parameterized reachability game problem, ranging from PTIME-complete to PSPACE-complete, in a variety of situations depending on the contraints (semilinear predicates, unions of intervals, or intervals) and on the presence or not of non-determinism.

7.4. New results on Axis 4: Data-driven Models

7.4.1. Crowdsourcing

Participants : Loïc Hélouët, Rituraj Singh

7.4.1.1. Complex workflows for crowdsourcing.

Crowdsourcing consists in hiring workers on internet to perform large amounts of simple, independent and replicated work units. We have proposed [32] complex workflows, a model for concurrent orcestration of tasks to solve problems that are more intricate than simpe tagging problems. Complex workflows allow higher-order answers where workers can suggest a process to obtain data rather than a plain answer. It is a data-centric model based on orchestration of concurrent tasks and higher order schemes. We have considered formal properties of specifications described with this model termination (whether some/all runs of a complex workflow terminate) and correctness (whether some/all runs of a workflow terminate with data satisfying FO requirements). We have shown that existential termination/correctness are undecidable in general excepted for specifications with bounded recursion. However, universal termination/correctness are decidable when constraints on inputs are specified in a decidable fragment of FO, and are at least in 2EXPTIME.

7.4.1.2. CrowdInc : a solution to reduce the cost of Consensus in Crowdsourcing.

Another contribution around crowdsourcing [34] considers agregation of answers, reliability of computed results, and optimization of costs. Crowdsoucing call for human expertise to solve problems which are still hard for computers, but easy for human workers. Crowdsourcing platform distribute replicated tasks to workers, pay them for their contribution, and aggregate answers to produce a reliable conclusion. A fundamental problem is to infer a correct answer from the set of results returned by workers. An additional ingredient of crowdsourcing is the cost needed to obtain a reliable answer: unlimited budget allows for the use of large pools of workers for each task, or experts to improve reliability of aggregated answers, but a limited budget forces to use resources at best to synthesize an reasonably reliable answer. We have focused on crowdsourcing of simple tasks with boolean answers. In this setting, we have first defined a probabilistic inference technique to agregate answers. This allows to consider difficulty of tasks and expertise of workers when building a conclusion. We have then proposed a greedy algorithm that reduces the cost (i.e. the number of workers hired by a platform) needed to reach a consensual answer. This algorithm considers difficulty of task, budget provided by client and total tasks to dynamically adapt threshold at each stage and makes locally optimal choice while preserving accuracy. Last, we have shown efficiency of our algorithm on several benchmarks, and compared its performance to existing solutions.

7.4.2. Guarded Attribute Grammars and Petri net synthesis

Participants : Adrian Puerto Aubel, Éric Badouel

7.4.2.1. Service-oriented programming

We addressed [30] the problem of component reuse for the design of user-centric distributed collaborative systems modelled by Guarded Attribute Grammars. Following the contract-based specification of components we develop an approach to an interface theory for the components of a collaborative system in three stages: we define a composition of interfaces that specifies how the component behaves with respect to its environement, we introduce an implementation order on interfaces and finally a residual operation on interfaces characterizing the systems that, when composed with a given component, can complement it in order to realize a global specification.

The visit of Joskel Ngoufo, a doctoral student at Yaoundé University, was the occasion to initiate a new implementation of the Guarded Attribute Grammars engine, in Racket language, a dialect of Lisp that allows metalanguage facilities and graphical interfaces to be processed more easily than in Haskell, the language chosen for the previous implementation.

7.4.2.2. Coordination of public debate.

Our research on data-centric collaborative systems has focused this year on the modelling of debates [28], with the aim of producing a tool that makes it possible to automatically conduct them, while managing relevant documents and analysing the respective positions of the different interventions from the point of view of argumentation theory. To this end, we are collaborating with Carlo Ferigato, a researcher at the JRC (C.E. Ispra, Italy), an institute for which we jointly produced a report covering an overview of the different theories developed around the subject, as well as the main tools proposing solutions to this problem. The aim of this collaboration is at understanding the basic principles and the computer programs apt to coordinate a public debate with an overall aim at giving the bases for designing such programs. Computer programs for the coordination of public debate exist since the beginning of the eighties but recently they have acquired new relevance for the use made of them by public administrations, associations and political parties. The meet of both citizen's needs and public administrations for transparency can today be technically realized with such programs through the present communication means in a more efficient way with respect to the first experiments dating now about forty years. This report aims at covering historical, technical and some theoretical aspects of the use of computers for the coordination of public debate.

7.4.2.3. Orthomodular partial orders.

The collaboration with Carlo Ferigato, is in line with the latter's thesis subject [11]. The set of regions of a condition/event transition system represents all the possible local states of a net system the behaviour of which is specified by the transition system. This set can be endowed with a structure, so as to form an orthomodular partial order. Given such a structure, one can then define another condition/event transition system. We study cases in which this second transition system has the same collection of regions as the first one. When it is so, the structure of regions is called stable. We proposed, to this aim, a composition operation, and a refinement operation for stable orthomodular partial orders, the results of which are stable.

7.5. New results on Transversal Concern: Missing Models

Participants : Hugo Bazille, Sihem Cherrared, Éric Fabre, Blaise Genest, Thierry Jéron, The Anh Pham

7.5.1. Unfolding-based dynamic partial-order reduction of asynchronous distributed programs

Unfolding-based Dynamic Partial Order Reduction (UDPOR) is a recent technique mixing Dynamic Partial Order Reduction (DPOR) with concepts of concurrency such as unfoldings to efficiently mitigate state space explosion in model-checking of concurrent programs. It is optimal in the sense that each Mazurkiewicz trace, *i.e.* a class of interleavings equivalent by commuting independent actions, is explored exactly once. In this work [25] we show that UDPOR can be extended to verify asynchronous distributed applications, where processes both communicate by messages and synchronize on shared resources. To do so, a general model of asynchronous distributed programs is formalized in TLA+. This allows to define an independence relation, a main ingredient of the unfolding semantics used by UDPOR during the UDPOR exploration. Then, the adaptation of UDPOR, involving the construction of an unfolding during the execution of the applicaton (*i.e.* with no model of the application but the code iteself), is made efficient by a precise analysis of dependencies. A prototype implementation gives promising experimental results.

7.5.2. Learning models for telecommunication management.

Model based methods have been recognised as the most appropriate approach to fault diagnosis in telecommunication networks, as they not only help in detecting and classifying failures, but is also provides useful explanations about the propagation of faults in such large distributed and concurrent systems. However, the bottleneck of these methods is of course the derivation and validation of a relevant model [8]. We have explored two techniques in this direction, based on fault/stress injection.

A first approach (collaboration Orange Labs) [33] consists in assembling generic components that would match the current (changing) topology of a software defined network. The model can then be validated by fault injection on a platform running the true VNF (virtual network functions) chains that are used in production. The second approach (collaboration Nokia Bell Labs) aims at detecting soft performance degradations, that would impact the quality of service, but not produce faults and alarms. Again, this can be achieved by stress injection at the level of VMs (virtual machines) in production software, and by collecting signature patterns under the form of statistical changes in the performance metrics collected on such systems.

7.5.3. Verification of deep neural networks.

Deep neural networks are as effective in their respective tasks as hardly understandable by a human. To use them in critical applications, not only they should be understood, they must be certified. We surveyed in [14] a large number of recent attempts to formally certify deep neural networks obtained by deep machine learning techniques. Most of the work currently focus on forward-propagating networks, and the problem of certifying their robustness.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Nokia Bell Labs - ADR SAPIENS

Several researchers of SUMO are involved in the joint research lab of Nokia Bell Labs France and Inria. We participate in the common research team SAPIENS (Smart Automated and Programmable Infrastructures for End-to-end Networks and Services), previously named “Softwarization of Everything.” This team involves several other Inria teams: Convecs, Diverse and Spades. 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 study control and diagnosis issues for such systems.

Two PhD students are involved in the project. Erij Elmajed (3rd year), on the topic of Diagnosis of virtualized and reconfigurable systems supervised by Éric Fabre and Armen Aghasaryan (Nokia Bell Labs). Abdul Majith (started in January 2019) on Controller Synthesis of Adaptive Systems, supervised by Hervé Marchand, Ocan Sankur and Dinh Thai Bui (Nokia Bell Labs).

8.1.2. Orange Labs

SUMO takes part in IOLab, the common lab of Orange Labs and Inria, dedicated to the design and management of Software Defined Networks. Our activities concern the diagnosis of malfunctions in virtualized multi-tenant networks.

This collaboration supports one Cifre PhD student, Sihem Cherrared (2nd year), supervised by Éric Fabre, Gregor Goessler (Inria Spades, Grenoble) and Sofiane Imadali (Orange Labs).

8.1.3. Alstom Transport - P22

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 P22 project between Alstom and Inria ended in 2018. However, our collaboration with Alstom Transport continues. One of the outcomes of this collaboration is the PhD defense of Karim Kecir in July 2019 [2].

8.1.4. Mitsubishi Electric Research Center Europe (MERCE)

Several researchers of SUMO are involved in a collaboration on the verification of real-time systems with the "Information and Network Systems (INS)" Team led by David Mentré of the "Communication & Information Systems (CIS)" Division of MERCE Rennes. The members of the team at MERCE work on different aspects of formal verification. Currently the SUMO team and MERCE jointly supervise a Cifre PhD student (Emily Clément) funded by MERCE since fall 2018; the thesis is about robustness of reachability in timed automata. Moreover Reiya Noguchi, a young engineer, member of MERCE, on leave of a Japanese operational division of Mitsubishi is also hosted and co-supervised by the SUMO team since the beginning of 2019, one day per week; we collaborate with him on the consistency of timed requirements.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Rennes Métropole: Allocation d'Installation Scientifique (AIS)

- Individual grant, led by Nicolas Markey

The objective of this project is to explore two research directions in the continuity of recent works: a truly quantitative theory of formal verification on the one hand, and the development of strategy-synthesis algorithms for modular systems on the other hand. It ended in June 2019.

9.2. National Initiatives

9.2.1. ANR TickTac: Efficient Techniques for Verification and Synthesis of Real-Time Systems (2019-2023)

- Link to website
- Led by Ocan Sankur (SUMO);
- SUMO participants: Emily Clément, Léo Henry, Thierry Jéron, Nicolas Markey, Victor Roussanaly, Ocan Sankur
- Partners: LSV (Cachan), ISIR (Paris), LaBRI (Bordeaux), LRDE (Paris), LIF (Marseille)

The aim of TickTac is to develop novel algorithms for the verification and synthesis of real-time systems using the timed automata formalism. One of the project's objectives is to develop an open-source and configurable model checker which will allow the community to compare algorithms. The algorithms and the tool will be used on a motion planning case study for robotics.

9.2.2. ANR HeadWork: Human-Centric Data-oriented WORKflows (2016-2020)

- Link to website
- Led by David Gross-Amblard (Université Rennes 1);
- Participants : Éric Badouel, Loïc Hélouët, Adrian Puerto Aubel, Rituraj Singh;
- 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.

9.2.3. IPL HAC-SPECIS: High-performance Application and Computers, Studying PErformance and Correctness In Simulation (2016-2020)

- Link to website
- Led by Arnaud Legrand (Inria Grenoble 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, 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 interested in the extension of the SimGrid programming model of MPI with synchronization primitives, the formalisation in ATL, of this model, and its adaptation to dynamic partial-order-reduction methods (DPOR) that allow to reduce the explored state space. A prototype implementation of an existing method that combines DPOR with true-concurrency models has been experimented on toy examples. The Anh Pham completed his PhD in december 2019.

9.2.4. National informal collaborations

The team collaborates with the following researchers:

- Béatrice Bérard (LIP6, Paris 6) on problems of opacity and diagnosis, and on problems related to logics and partial orders for security;
- Patricia Bouyer (LSV, ENS Paris-Saclay) on the analysis of probabilistic timed systems and quantitative aspects of verification;
- Thomas Chatain and Stefan Haar (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;
- Gwenaël Delaval and Éric Rutten (Inria team Ctrl-A, LIG, Université Grenoble-Alpes) on the control of reconfigurable systems and the link between Reax and Heptagon/BZR (http://bzr.inria.fr/);
- Serge Haddad (Inria team MExICo, LSV, ENS Paris-Saclay) on opacity and diagnosis;
- Loïg Jézéquel (LS2N, Université de Nantes) on stochastic and timed nets, and on distributed optimal planning;
- Didier Lime and Olivier H. Roux (LS2N, Université de Nantes) on stochastic and timed Petri nets;
- François Laroussinie (IRIF, UP7-Diderot) on logics for multi-agent systems,

9.3. International Initiatives

9.3.1. Inria International Labs

LIRIMA: International Laboratory for Research in Computer Science and Applied Mathematics

9.3.1.1. FUCHSIA

Associate Team involved in the international lab LIRIMA.

Title: Flexible user-centric higher-order systems for collective intelligence in agencies

International Partner

U. Yaoundé (Cameroon) Georges-Edouard Kouamou

Start year: 2019

See also: https://project.inria.fr/fuchsia/

Develop methods and tools, based on guarded attribute grammars, to design flexible and adaptive systems for information gathering and deliberation in order to collaboratively build expertise in health emergency situations.

9.3.2. Inria Associate Teams Not Involved in an Inria International Labs

9.3.2.1. EQUAVE

Title: Efficient Quantitative Verification

International Partner

Indian Institute of Technology Bombay (India) - Dpt of Computer Science and Engineering - S. Akshay

Start year: 2018

See also: http://www.irisa.fr/sumo/EQUAVE

Formal verification has been addressed for a long time. A lot of effort has been devoted to Boolean verification, i.e., formal analyis of systems that check whether a given property is true or false.

In many settings, a Boolean verdict is not sufficient. The notions of interest are for instance the amount of confidential information leaked by a system, the proportion of some protein after a duration in some experiment in a biological system, whether a distributed protocol satisfies some property only for a bounded number of participants... This calls for quantitative verification, in which algorithms compute a value such as the probability for a property to hold, the mean cost of runs satisfying it, the time needed to achieve a complex workflow...

A second limitation of formal verification is the efficiency of algorithms. Even for simple questions, verification is rapidly PSPACE-complete. However, some classes of models allow polynomial time verification. The key techniques to master complexity are to use concurrency, approximation, etc

The objective of this project is to study efficient techniques for quantitative verification, and develop efficient algorithms for models such as stochastic games, timed and concurrent systems.

9.3.3. Inria International Partners

9.3.3.1. Informal International Partners

The team collaborates with the following researchers:

- S. Akshay (IIT Bombay, India) on timed concurrent models;
- Andrea D'Ariano (University Roma Tre, Italy), on train regulation.
- Christel Baier (Technical University of Dresden, Germany) on verification and control of stochastic systems;
- Thomas Brihaye (Université de Mons, Belgium) on the verification of stochastic timed systems;
- Gilles Geeraerts and Jean-François Raskin, (Université Libre de Bruxelles, Belgium) on multiplayer game theory and synthesis;
- Alessandro Giua and Michele Pinna (University Cagliari, Italy) on diagnosis and unfolding techniques for concurrent systems.

- Igor Konnov (Interchain, Austria), Marijana Laźic (Technical University Munich, Germany) and Josef Widder (Interchain, Austria) on the automated verification of randomized distributed algorithms.
- Stéfane Lafortune (University of Michigan, USA) on the control of cyber-physical systems;
- Kim G. Larsen (University Aalborg, Denmark) on quantitative timed games, and on topics related to urban train systems modeling;
- John Mullins (Polytechnique Montréal, Canada) on security and opacity;
- Mickael Randour (Université de Mons, Belgium) on quantitative games for synthesis.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- S. Akshay (IIT Bombay, India) visited the team for one week.
- Christel Baier and Jakob Piribauer (TU Dresden, Germany) visited the SUMO team for one week in september.
- Khushraj Nanik Madnani (IIT Bombay, India) visited our team during two months.
- Laurie Ricker (Mount Allison University, Canada) visited the team during 2 months.
- Graeme Zinck (Mount Allison University, Canada) visited our team during four months. He obtained a 5000$ grant provided by Mitacs through a collaboration between Mount Allison University (L. Ricker) and Inria (Loïc Hélouët and Hervé Marchand). Two papers are in preparation (one regarding the enforcement of opacity for modular systems (submitted to Ifac World congress) and the other about the enforcement of concurrent secrets for multiple systems.

9.4.1.1. Internships

- Pierre Boudart, ENS Ulm, June-July 2019, Éric Fabre.
- Kritin Garg and Sharvik Mital, IIT Bombay, May-July 2019, Éric Fabre, Blaise Genest and Loïc Hélouët.
- Mathieu Poirier, ENS Rennes, May-July 2019, Éric Badouel and Adrian Puerto Aubel.
- Bastien Thomas, ENS Rennes, Feb-July 2019, Nathalie Bertrand.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

10.1.1.1. General Chair, Scientific Chair

- Hervé Marchand is a member of the IFAC Technical Committees (TC 1.3 on Discrete Event and Hybrid Systems). He is the president of the steering committee of MSR (modélisation de systèmes réactifs).
- Nathalie Bertrand and Nicolas Markey are members of the steering committee of the Summer School MOVEP (*Modélisation et Vérification des Processus Parallèles*).
- Blaise Genest is member of the steering comittee of the international workshop FMAI (*Formal Methods and Artificial Intelligence*).

10.1.1.2. Member of the Organizing Committees

- Blaise Genest coorganized the 2nd workshop FMAI 2019 (Rennes, 2-3 May 2019).

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

- Éric Badouel was member of the Program Committees of VECOS, ATAED, ICTAC, CRI, and JIMIS in 2019.
- Nathalie Bertrand was a member of the Program Committees of the following international events: TIME'19, RP'19, MFCS'19.
- Éric Fabre was a member of the Program Committee of CREST'19.
- Blaise Genest was a member of the Program Committee of FMAI'19.
- Loïc Hélouët was member of the Program Committee of ACSD'2019.
- Thierry Jéron served on the Program Committees of ICTSS'19 and SAC-SVT'20.
- Nicolas Markey was a member of the Program Committee of LATA'20.
- Ocan Sankur was a member of the Program Committee of FORMATS'19.

10.1.2.2. Reviewer

In 2019, members of SUMO reviewed submissions for following conferences: VECOS, ATAED, CARI, ICTAC, CONCUR, SOFSEM, FOCS, ATVA, VMCAI, ICALP, SAC-SVT, TAP, ACSD, MFCS, STACS, WODES, HSCC, FSTTCS, CSL, AAMAS, TACAS, FoSSaCS, LICS, PODC, MORE, RP.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Éric Badouel is co-editor-in-Chief of ARIMA Journal.
- Hervé Marchand is associate editor of the journal Discrete Event Dynamical Systems - Theory and applications since january 2019.

10.1.3.2. Reviewer - Reviewing Activities

In 2019, members of SUMO reviewed submissions for following journals: Automatica, Fundamenta Informaticae, Information and Computation, The Scientific Annals of Computer Science, Science of Computer Programming, ACM Transactions on Computational Logic, ACM Transactions on Embedded Computing Systems, Journal of Systems and Software, Mathematical Review (MathSciNet), Journal of Discrete Event Dynamical Systems, Formal Methods in System Design, Software Testing, Verification and Reliability, Journal of Logic and Computation, IEEE Transactions on Automatic Control, PLoS one, Performance Evaluation, Artificial Intelligence, Journal of Logic and Algebraic Methods in Programming, Logical Methods in Computer Science, ACM Transactions on Modeling and Computer Simulation, Journal of Systems and Software.

10.1.4. Invited Talks

- Nathalie Bertrand gave an invited talk at the international conference Formats'19 on Taming real-time stochastic systems.
- Blaise Genest was invited to the workshop SinFra'19 in Singapore and gave a talk on Trust in AI.
- Léo Henry was invited to a workshop on test generation by IMDEA (Madrid) to give a talk about test generation for timed automata using games.

10.1.5. Leadership within the Scientific Community

Nathalie Bertrand is the co-head of the *Groupe de Travail Vérif* (together with Pierre-Alain Reynier (LIS, Marseille)) which is part of *GDR Informatique Mathématique (GDR-IM)*.

10.1.6. Scientific Expertise

- Nathalie Bertrand was a reviewer for Thelam Fund and FWO (Belgium) and Grenoble-MSTIC. She served on the HCERES committee for Vérimag.
- Blaise Genest was reviewer for a DIGICOSME project.
- Loïc Hélouët was reviewer for ANR.
- Thierry Jéron was a reviewer for NWO (Netherlands Organisation for Scientific Research).
- Nicolas Markey served on the HCERES committee for LACL.

10.1.7. 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 DPEI (European and International Partnership Department). He is member of the executive board of GIS SARIMA.
- Nathalie Bertrand was nominated member of the Conseil National des Universités, section 27 (computer science) until November 2019.
- Emily Clément is a representative of PhD students in the Comité de Centre of Inria Rennes.
- 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 started in 2017, 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 SAPIENS.
  Éric Fabre is also a member of Inria Evaluation Commission since September 2019.
- Loïc Hélouët is member of Inria CNHSCT (committee for Health and Security). He is also a suppletive member in the Comité de Centre of Inria Rennes. He leads a working group of the comittee on harrassment, and another of daily life improvement. He is member of a commission at IRISA on harrassment.
- Thierry Jéron is a member of the IFIP Working Group 10.2 on Embedded Systems. He is a member of the Comité d'orientation scientifique (COS) of IRISA Rennes. He was member of the Comité de Centre of Inria Rennes until mid-2019. Since 2016 he is "référent chercheur" for the Inria-Rennes research center.
- Hervé Marchand was chairman of the *Comission des utilisateurs des moyens informatiques* (CUMI) in Rennes until December 2019. He is an elected member of the Comité de Centre at Inria Rennes since June 2019.
- Nicolas Markey manages the mentoring programme at Irisa/InriaRBA; this programme aims at having senior researchers transfering their experience to younger colleagues (including PhD students and postdoc). The programme currently concerns about 30 mentor/mentee pairs.

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 algorithms, L2, 40h, INSA de Rennes, France.

Master: Éric Badouel, Logic and argumentation, 32h, Univ Yaoundé I, Cameroon.

Master: Nathalie Bertrand, Language Theory; Algorithms, 20h, 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: Loïc Hélouët, Algorithms, 4h, Agrégation, ENS Rennes, France;
Master: Loïc Hélouët, Algorithms and proof, 12h, Agrégation, ENS Rennes, France;
Master: Nicolas Markey, Verification of Complex Systems (CSV), 15h, M2, Univ Rennes 1, France;
Master: Nicolas Markey, Algorithms, 12h, Agrégation, ENS Rennes, France;
Master: Ocan Sankur, Verification of Complex Systems (CSV), 10h, M2, Univ Rennes 1, France;
Master: Ocan Sankur, *Travaux pratiques*, Analyse et Conception Formelle (ACF), 22h, M1, Univ Rennes 1, France;

10.2.2. Supervision

10.2.2.1. PhD Students

- PhD: Robert Fondze Jr Nsaibirni, A Guarded Attribute Grammar Based Model for User Centered, Distributed, and Collaborative Case Management – Case of the Disease Surveillance Process [3], supervised by Éric Badouel. Defended at the University of Yaoundé I, Cameroon in April 2019.
- PhD: Karim Kecir, Performance Evaluation of Urban Rail Traffic Management Techniques [2], supervised by Loïc Hélouët and Pierre Dersin (Alstom), Université Rennes 1, July 2019.
- PhD: Samy Jaziri, Automata on Timed Structures, supervised by Nicolas Markey, Université Paris Saclay, September 2019.
- PhD: Mauricio Gonzalez, Stochastic Games on Graphs with Applications to Smart-Grids Optimization, supervised by Nicolas Markey, Université Paris Saclay, November 2019.
- PhD: Hugo Bazille, Detection and Quantification of Events in Stochastic Systems [1], supervised by Blaise Genest and Éric Fabre, Université Rennes 1, December 2019.
- PhD: The Anh Pham, Efficient state-space exploration for asynchronous distributed programs - Adapting unfolding-based dynamic partial order reduction to MPI programs [4], supervised by Thierry Jéron and Martin Quinson (Myriads, Inria Rennes), ENS Rennes, December 2019.

- PhD in progress: Sihem Cherrared, Diagnosis of multi-tenant programmable networks, started Dec. 2016, Éric Fabre, Gregor Goessler (Inria, Spades) and Sofiane Imadali (Orange).
- PhD in progress: Emily Clément, Verification and synthesis of control systems: efficiency and robustnes, started Dec. 2018, supervised by Thierry Jéron, Nicolas Markey, and David Mentré (Mitsubishi Electric)
- PhD in progress: Rodrigue Djeumen Djatcha, Collaborative Model for Urban Crowdsourcing, started in September 2017, University of Douala, Cameroon, supervised by Éric Badouel.
- PhD in progress: Erij Elmajed, Diagnosis of reconfigurable systems, started March 2017, Éric Fabre and Armen Aghasaryan (Nokia).
- PhD in progress: Léo Henry, Optimal test-case generation with game theory, started Oct. 2018, supervised by Thierry Jéron and Nicolas Markey.
- PhD in progress: Abdul Majith, Control of Adaptive Systems, started in Jan. 2019, supervised by Hervé Marchand, Ocan Sankur, and Dinh Thai-Bui (Nokia Bell Labs).
- PhD in progress: Anirban Majumdar, Games for distributed networks: models and algorithms, ENS Paris Saclay, started Sept 2018, supervised by Nathalie Bertrand and Patricia Bouyer (LSV).
- PhD in progress: Arthur Queffelec, Tradeoff between Robustness and Optimality in Strategic Reasoning, started Nov. 2018, supervised by Ocan Sankur and François Schwarzentruber (Logica, IRISA).
- PhD in progress: Victor Roussanaly, Efficient verification of timed systems, started Sep. 2017, supervised by Nicolas Markey and Ocan Sankur.
- PhD in progress: Suman Sadhukhan, Modelling and parameterized verification of mobile networks, started Oct. 2018, supervised by Nathalie Bertrand, Nicolas Markey and Ocan Sankur.
- PhD in progress: Rituraj Singh, Data-centric Workflows for Crowdsourcing Applications, started Feb. 2018, supervised by Loïc Hélouët.
- PhD in progress: Bastien Thomas, Automated verification of randomized distributed algorithms, started in Oct. 2019, supervised by Nathalie Bertrand and Josef Widder (Interchain, Austria).

10.2.2.2. Master Students

- Nathalie Bertrand supervised the master's thesis (M2) of Bastien Thomas, feb-june 2019.
- Blaise Genest and and Léo Henry supervise (2 h/week during 6 months) Alexandre Drewery, a master 1 student. The topic is reinforcement learning of mixed discrete/continuous systems.

10.2.2.3. Other Internships

- L3 Internship of Pierre Boudard, ENS ULM, supervised by Éric Fabre.
- L2 Internship of Kritin Garg, supervised by Éric Fabre and Blaise Genest.
- L3 Internship of Sharvik Mital, supervised by Blaise Genest and Loïc Hélouët.
- L3 Internship of Mathieu Poirier, supervised by Éric Badouel and Adrian Puerto Aubel.

10.2.3. Juries

10.2.3.1. PhD Defenses

- Nathalie Bertrand was an examiner for the PhD thesis of Damien Busatto-Gaston (Université Aix-Marseille, december 2019).
- Éric Fabre took part to the jury for the PhD in Computer Science of Maha Mdini, Institut Mines Telecom (IMT) Atlantique, Sept. 2019
- Blaise Genest was a reviewer for the PhD of Sukanya Basu, IIT Bombay, India.
- Hervé Marchand was an examiner in the PhD defense of Raphael Jakse, Université Grenoble Alpes in December 2019.
- Nicolas Markey was a reviewer for the PhD thesis of Nicola Gigante, Jan 2019, University Udine, Italy.

10.2.3.2. Other Juries

- Nathalie Bertrand was in the *Moyens incitatifs* committee for Inria Rennes Bretagne Atlantique in 2019.
- Éric Fabre was in the hiring committee for CRCN positions at Inria Rennes Bretagne Atlantique in 2019.
- Ocan Sankur was in the hiring committee for two Maitre de conférences positions at Université de Nantes in 2019.

10.3. Popularization

10.3.1. Education

Nicolas Markey is involved in the organization of action "J'Peux Pas, J'Ai Informatique", whose aim is to break down stereotypes about computer science for 12-year-old pupils.

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] H. BAZILLE.*Detection and Quantification of Events in Stochastic Systems*, Univ-Rennes1, December 2019, https://hal.inria.fr/tel-02415750

[2] A.-E.-K. KECIR.*Performance evaluation of urban rail traffic management techniques*, Université Rennes 1, July 2019, https://tel.archives-ouvertes.fr/tel-02317224

[3] R. NSAIBIRNI.*A Guarded Attribute Grammar Based Model for User Centered, Distributed, and Collaborative Case Management Case of the Disease Surveillance Process*, Université de Yaoundé I, April 2019, https://hal.inria.fr/tel-02263094

[4] T. A. PHAM.*Efficient state-space exploration for asynchronous distributed programs: Adapting unfolding-based dynamic partial order reduction to MPI programs*, ENS Rennes, December 2019, https://hal.archives-ouvertes.fr/tel-02420950

Articles in International Peer-Reviewed Journal

[5] O. BEAUMONT, T. LAMBERT, L. MARCHAL, B. THOMAS.*Performance Analysis and Optimality Results for Data-Locality Aware Tasks Scheduling with Replicated Inputs*, in "Future Generation Computer Systems", October 2019, p. 1-28 [*DOI :* 10.1016/J.FUTURE.2019.08.024], https://hal.inria.fr/hal-02275473

[6] N. BERTRAND, M. DEWASKAR, B. GENEST, H. GIMBERT, A. GODBOLE.*Controlling a population*, in "Logical Methods in Computer Science", 2019, vol. 15, n^o 3, p. 1-30 [*DOI :* 10.23638/LMCS-15(3:6)2019], https://hal.archives-ouvertes.fr/hal-02350251

[7] N. BERTRAND, S. HADDAD, E. LEFAUCHEUX.*A Tale of Two Diagnoses in Probabilistic Systems*, in "Journal of Information and Computation", December 2019, vol. 269, p. 1-33 [*DOI :* 10.1016/J.IC.2019.104441], https://hal.inria.fr/hal-02430814

[8] S. CHERRARED, S. IMADALI, E. FABRE, G. GÖSSLER, I. G. B. YAHIA.*A Survey of Fault Management in Network Virtualization Environments: Challenges and Solutions*, in "IEEE Transactions on Network and Service Management", October 2019, p. 1-15 [*DOI :* 10.1109/TNSM.2019.2948420], https://hal.inria.fr/hal-02370378

[9] M. PICHENÉ, S. K. PALANIAPPAN, E. FABRE, B. GENEST.*Modeling Variability in Populations of Cells using Approximated Multivariate Distributions*, in "IEEE/ACM Transactions on Computational Biology and Bioinformatics", 2019, p. 1-12, forthcoming [*DOI :* 10.1109/TCBB.2019.2904276], https://hal.archives-ouvertes.fr/hal-02350249

[10] M. RENARD, Y. FALCONE, A. ROLLET, T. JÉRON, H. MARCHAND.*Optimal Enforcement of (Timed) Properties with Uncontrollable Events*, in "Mathematical Structures in Computer Science", 2019, vol. 29, n^o 1, p. 169-214 [*DOI :* 10.1017/S0960129517000123], https://hal.archives-ouvertes.fr/hal-01262444

International Conferences with Proceedings

[11] F. ADOBBATI, C. FERIGATO, S. GANDELLI, A. PUERTO AUBEL.*Two Operations for Stable Structures of Elementary Regions*, in "ATAED 2019 - Workshop Algorithms & Theories for the Analysis of Event Data", Aachen, Germany, W. VAN DER AALST, R. BERGENTHUM, J. CARMONA (editors), Proceedings of the International Workshop on Algorithms & Theories for the Analysis of Event Data 2019, CEUR Workshop Proceedings, June 2019, vol. 2371, n^o 3, p. 36-53, https://hal.inria.fr/hal-02337628

[12] S. AKSHAY, H. BAZILLE, E. FABRE, B. GENEST.*Classification among Hidden Markov Models*, in "FSTTCS 2019 - 39th IARCS Annual Conference on. Foundations of Software Technology and Theoretical Computer Science", Bombay, India, Proceedings of FSTTCS 2019 - 39th IARCS Annual Conference on. Foundations of Software Technology and Theoretical Computer Science, LIPIcs, 2019, vol. volume 150 of LIPIcs, n^o 29, p. 1-14 [*DOI :* 10.4230/LIPICS.FSTTCS.2019.29], https://hal.archives-ouvertes.fr/hal-02350252

[13] C. BAIER, N. BERTRAND, J. PIRIBAUER, O. SANKUR.*Long-run Satisfaction of Path Properties*, in "LICS 2019 - 34th Annual ACM/IEEE Symposium on Logic in Computer Science", Vancouver, Canada, IEEE, June 2019, p. 1-31 [*DOI :* 10.1109/LICS.2019.8785672], https://hal.archives-ouvertes.fr/hal-02349456

[14] H. BAZILLE, E. FABRE, B. GENEST.*Certification formelle des réseaux neuronaux profonds : un état de l'art en 2019*, in "AI and Defense 2019 - Artificial Intelligence and defense", Rennes, France, Actes de AI&Defense, 2019, p. 1-10, https://hal.archives-ouvertes.fr/hal-02350253

[15] N. BERTRAND, B. BORDAIS, L. HÉLOUËT, T. MARI, J. PARREAUX, O. SANKUR.*Performance Evaluation of Metro Regulations Using Probabilistic Model-checking*, in "RSSRail 2019 - International conference on reliability, safety and security of railway systems: modelling, analysis, verification and certification", Lille, France, LNCS, Springer, June 2019, p. 59-76 [*DOI :* 10.1007/978-3-030-18744-6_4], https://hal.inria.fr/hal-02065365

[16] N. BERTRAND, P. BOUYER, A. MAJUMDAR.*Concurrent parameterized games*, in "FSTTCS 2019 - 39th IARCS Annual Conference on. Foundations of Software Technology and Theoretical Computer Science", Bombay, India, LIPIcs, December 2019, p. 1-15, https://hal.inria.fr/hal-02351236

[17] N. BERTRAND, P. BOUYER, A. MAJUMDAR.*Reconfiguration and message losses in parameterized broadcast networks*, in "CONCUR 2019 - 30th International Conference on Concurrency Theory", Amsterdam, Netherlands, August 2019, p. 1 - 15 [*DOI :* 10.4230/LIPICS.CONCUR.2019.32], https://hal.inria.fr/hal-02191382

[18] N. BERTRAND, I. KONNOV, M. LAZIC, J. WIDDER.*Verification of Randomized Consensus Algorithms under Round-Rigid Adversaries*, in "CONCUR 2019 - 30th International Conference on Concurrency Theory", Amsterdam, Netherlands, August 2019, p. 1-16 [*DOI :* 10.4230/LIPICS.CONCUR.2019.33], https://hal.inria.fr/hal-02191348

[19] P. BOUYER, O. KUPFERMAN, N. MARKEY, B. MAUBERT, A. MURANO, G. PERELLI.*Reasoning about Quality and Fuzziness of Strategic Behaviours*, in "IJCAI 2019 - 28th International Joint Conference on Artificial Intelligence", Macao, China, Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI'19), August 2019, p. 1588-1594, https://arxiv.org/abs/1905.11537 [*DOI :* 10.24963/IJCAI.2019/220], https://hal.archives-ouvertes.fr/hal-02268141

[20] D. BUSATTO-GASTON, B. MONMEGE, P.-A. REYNIER, O. SANKUR.*Robust Controller Synthesis in Timed Büchi Automata: A Symbolic Approach*, in "CAV 2019 - 31st International Conference on Computer Aided Verification", New-York, United States, LNCS, Springer, July 2019, p. 572-590 [*DOI :* 10.1007/978-3-030-25540-4_33], https://hal.archives-ouvertes.fr/hal-02264083

[21] T. CHARRIER, A. QUEFFELEC, O. SANKUR, F. SCHWARZENTRUBER.*Reachability and Coverage Planning for Connected Agents*, in "AAMAS 2019 - 18th International Conference on Autonomous Agents and MultiAgent Systems", Montreal, Canada, Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, ACM, May 2019, vol. 3, p. 1874-1876, https://hal.archives-ouvertes.fr/hal-02349490

[22] T. CHARRIER, A. QUEFFELEC, O. SANKUR, F. SCHWARZENTRUBER.*Reachability and Coverage Planning for Connected Agents*, in "IJCAI 2019 - 28th International Joint Conference on Artificial Intelligence", Macao, China, August 2019, p. 1-7, https://hal.archives-ouvertes.fr/hal-02349475

[23] L. HÉLOUËT, N. MARKEY, R. RAHA.*Reachability Games with Relaxed Energy Constraints*, in "GandALF 2019 - Tenth International Symposium on Games, Automata, Logics, and Formal Verification", Bordeaux, France, EPTCS, September 2019, vol. 305, p. 17-33 [*DOI :* 10.4204/EPTCS.305.2], https://hal.inria.fr/hal-02291241

[24] R. MEIRA-GÓES, H. MARCHAND, S. LAFORTUNE.*Towards resilient supervisors against sensor deception attacks*, in "CDC 2019 - 58th IEEE Conference on Decision and Control", Nice, France, IEEE, December 2019, p. 1-6, https://hal.inria.fr/hal-02390435

[25] T. A. PHAM, T. JÉRON, M. QUINSON.*Unfolding-based Dynamic Partial Order Reduction of Asynchronous Distributed Programs*, in "FORTE 2019 - 39th International Conference on Formal Techniques for Distributed Objects, Components, and Systems", Copenhagen, Denmark, J. A. PÉREZ, N. YOSHIDA (editors), Formal Techniques for Distributed Objects, Components, and Systems, Springer International Publishing, 2019, vol. LNCS-11535, p. 224-241, Part 1: Full Papers [*DOI :* 10.1007/978-3-030-21759-4_13], https://hal.inria.fr/hal-02109769

[26] V. ROUSSANALY, O. SANKUR, N. MARKEY.*Abstraction Refinement Algorithms for Timed Automata*, in "CAV 2019 - 31st International Conference on Computer Aided Verification", New York, United States, LNCS, Springer, July 2019, vol. 11561, p. 22-40 [*DOI :* 10.1007/978-3-030-25540-4_2], https://hal.archives-ouvertes.fr/hal-02265808

Conferences without Proceedings

[27] M. GONZALEZ, P. BOUYER, S. LASAULCE, N. MARKEY.*Optimisation en présence de contraintes en probabilité et processus markoviens contrôlés*, in "GRETSI 2019 - XXVIIème Colloque GRETSI Traitement du Signal & des Images", Lille, France, August 2019, p. 1-4, https://hal.archives-ouvertes.fr/hal-02268161

Research Reports

[28] E. BADOUEL, C. FERIGATO, P. AUBEL.*Computers and Coordination of Debate. A study on the role of computers for ordering public debates at various levels, from open citizen's polls to formal parliamentary debate*, Euopean Community Joint Research Center - Ispra ; Inria Rennes - Bretagne Atlantique ; Université Rennes 1 ; Irisa, November 2019, n^{o} JRC115574, 63, https://hal.inria.fr/hal-02346119

Other Publications

[29] S. AKSHAY, B. GENEST, L. HÉLOUËT, S. MITAL.*Timed Negotiations*, October 2019, working paper or preprint, https://hal.inria.fr/hal-02337887

[30] E. BADOUEL, R. A. DJEUMEN DJATCHA.*A Calculus of Interfaces for Guarded Attribute Grammars*, June 2019, working paper or preprint, https://hal.inria.fr/hal-02145920

[31] N. BERTRAND, I. KONNOV, M. LAZIC, J. WIDDER.*Verification of Randomized Distributed Algorithms under Round-Rigid Adversaries*, April 2019, Experiments presented in this paper were carried out using the Grid5000 testbed, supported by a scientific interest group hosted by Inria and including CNRS, RENATER and several Universities as well as other organizations, see grid5000.fr, https://hal.inria.fr/hal-01925533

[32] P. BOURHIS, L. HÉLOUËT, R. SINGH, Z. MIKLÓS.*Data Centric Workflows for Crowdsourcing*, September 2019, working paper or preprint, https://hal.inria.fr/hal-01976280

[33] S. CHERRARED, S. IMADALI, E. FABRE, G. GÖSSLER.*SAKURA a Model Based Root Cause Analysis Framework for vIMS*, ACM Press, June 2019, p. 594-595, MobiSys 2019 - 17th ACM International Conference on Mobile Systems, Applications, and Services, Poster, https://hal.inria.fr/hal-02291163

[34] R. SINGH, L. HÉLOUËT, Z. MIKLÓS.*Reducing the Cost of Aggregation in Crowdsourcing*, December 2019, working paper or preprint, https://hal.inria.fr/hal-02397971

References in notes

[35] E. BADOUEL, U. SCHLACHTER.*Incremental Process Discovery using Petri Net Synthesis*, in "Fundamenta Informaticae", June 2017, vol. 154, n° 1-4, p. 1-13 [*DOI :* 10.3233/FI-2017-1548], https://hal.inria.fr/hal-01599760

[36] N. BERTRAND, P. BOUYER, T. BRIHAYE, P. CARLIER.*Analysing Decisive Stochastic Processes*, in "ICALP 2016 - 43rd International Colloquium on Automata, Languages, and Programming", Rome, Italy, LiPIcs, LZI, 2016, vol. 55, p. 101:1-101:14 [*DOI :* 10.4230/LIPICS.ICALP.2016.101], https://hal.inria.fr/hal-01397794

[37] N. BERTRAND, P. BOUYER, T. BRIHAYE, P. CARLIER.*When are stochastic transition systems tameable?*, in "Journal of Logical and Algebraic Methods in Programming", 2018, vol. 99, p. 41-96 [*DOI :* 10.1016/J.JLAMP.2018.03.004], https://hal.inria.fr/hal-01938135

[38] P. BOUYER, N. MARKEY, N. PERRIN, P. SCHLEHUBER-CAISSIER.*Timed automata abstraction of switched dynamical systems using control funnels*, in "Real-Time Systems", May 2017, vol. 53, n° 3, p. 327-353, http://dx.doi.org/10.1007/s11241-016-9262-3

[39] I. V. KONNOV, M. LAZIC, H. VEITH, J. WIDDER.*A short counterexample property for safety and liveness verification of fault-tolerant distributed algorithms*, in "POPL 2017 - 44th ACM SIGPLAN Symposium on Principles of Programming Languages", ACM, 2017, p. 719-734

[40] A. NIGAM, N. S. CASWELL.*Business artifacts: An approach to operational specification*, in "IBM Systems Journal", 2003, vol. 42, n° 3, p. 428-445

Project-Team TAMIS

Threat Analysis and Mitigation for Information Security

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

IN PARTNERSHIP WITH:
CNRS

CentraleSupélec

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Security and Confidentiality

Table of contents

Project-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
A4.3. - Cryptography
A4.4. - Security of equipment and software
A4.5. - Formal methods for security

Other Research Topics and Application Domains:

B6.6. - Embedded systems

1. Team, Visitors, External Collaborators

Research Scientists

Olivier Zendra [Team leader, Inria, Researcher]
Annelie Heuser [CNRS, Researcher]
Kim Larsen [Inria, International Chair, Advanced Research Position]

Technical Staff

Najah Ben Said [Inria, Engineer, until Jan 2019]
Ioana Domnina Cristescu [Inria, Engineer, until Jul 2019]
Cassius de Oliveira Puodzius [Inria, Engineer, until Jan 2019]
Olivier Decourbe [Inria, Engineer]
Nicolas Kiss [Inria, Engineer, from Apr 2019]
Bruno Lebon [Inria, Engineer]
Céline Minh [Inria, Engineer]

PhD Students

Delphine Beaulaton [UBS Vannes, PhD Student, until Jul 2019]
Cassius de Oliveira Puodzius [Inria, PhD Student, from Feb 2019]
Christophe Genevey-Metat [Univ de Rennes I, PhD Student]
Alexandre Gonzalvez [IMT Atlantique, PhD Student, until Nov 2019]
Nisrine Jafri [Inria, PhD Student, until Feb 2019]
Tristan Ninet [Thales, PhD Student, until Nov 2019, granted by CIFRE]
Lamine Noureddine [Inria, PhD Student]
Duy Phuc Pham [CNRS, PhD Student, from May 2019]
Alexander Zhdanov [Inria, PhD Student, until Jun 2019]

Post-Doctoral Fellows

Eduard Baranov [Inria, Post-Doctoral Fellow, until Feb 2019]
Yoann Marquer [Inria, Post-Doctoral Fellow]
Matthieu Mastio [CNRS, Post-Doctoral Fellow, from Mar 2019]
Tania Richmond [Inria, Post-Doctoral Fellow]
Stefano Sebastio [Inria, Post-Doctoral Fellow]

Visiting Scientist

Nisrine Jafri [Inria, from Mar 2019 until Apr 2019]

Administrative Assistant

Cécile Bouton [Inria, Administrative Assistant]

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 [18].

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.

4. Highlights of the Year

4.1. Highlights of the Year

4.1.1. Kick-off of the ANR JCJC AHMA project

The ANR JCJC project lead by Annelie Heuser was kicked-off, and a PostDoc (Matthieu Mastio) and PhD (Duy Phuc Pham) have been hired. The team already created a first platform for automated hardware malware analysis. See below and in the following.

4.1.2. New results in the TeamPlay H2020 project, coordinator

The project is coordinated by Olivier Zendra. The TeamPlay H2020 project had a successful mid-term review in October 2019, where the reviewers stressed the quality of the overall work. We TAMIS also achieved new results on security modelling in this TeamPlay project in 2019 (see in the following).

4.1.3. New software and platforms

In 2019, we continued the development of several software and platforms (hardware and software), and build up four new ones:

- E-PAC, an Evolving Packer Classifier,
- The SABR (Semantic-driven Analysis of BinaRies) platform
- Orqal, an efficient schedueler for docker images.
- A Side-channel deep learning evaluation platform,
- The AHMA (IoT malware classification through side-channel information) platform and tools.

5. New Software and Platforms

5.1. 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.

- Participants: Bruno Lebon, Olivier Zendra, Alexander Zhdanov and Fabrizio Biondi
- Contact: Bruno Lebon

5.2. IoTMLT

IoT Modeling Language and tool

KEYWORDS: Internet of things - Modeling language - Cyber attack

SCIENTIFIC DESCRIPTION: We propose a framework to analyze security in IoT systems consisting of a formal languages for modeling IoT systems and of attack trees for modeling the possible attacks on the system. In our approach a malicious entity is present in the system, called the Attacker. The other IoT entities can inadvertently help the Attacker, by leaking their sensitive data. Equipped with the acquired knowledge the Attacker can then communicate with the IoT entities undetected. The attack tree provided with the model acts as a monitor: It observes the interactions the Attacker has with the system and detects when an attack is successful.

An IoT system is then analyzed using statistical model checking (SMC). The first method we use is Monte Carlo, which consists of sampling the executions of an IoT system and computing the probability of a successful attack based on the number of executions for which the attack was successful. However, the evaluation may be difficult if a successful attack is rare. We therefore propose a second SMC method, developed for rare events, called importance splitting. Both methods are proposed by Plasma, the SMC tool we use.

FUNCTIONAL DESCRIPTION: The IoT modeling language is a formal language and tool for specifying and enforcing security in IoT systems.

- Participants: Delphine Beaulaton, Ioana-Domnina Cristescu and Najah Ben Said
- Partner: Vérimag
- Contact: Delphine Beaulaton
- URL: http://iot-modeling.gforge.inria.fr

5.3. 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

5.4. AHMA

Automatic Malware Hardware Analysis

KEYWORDS: Side-channel - Deep learning - Malware

FUNCTIONAL DESCRIPTION: This framework is composed of several parts, each one of them taking in charge the generation and the processing of the data at different levels. Drivers have been developed to automatically control the different oscilloscopes we are working with (picoscope 6407 et infiniium keysight). We use signal processing tools on the raw data to feed a deep neural network which is in charge of classifying the observed malwares. We are using two different approaches to manage the infection of the system. The first one is to reinitialize it each time we make a measurement to ensure its integrity. We have proposed a method allowing to speed the procedure up a lot. Besides, we developed several malwares, to make our experiments in a controlled environment, to avoid the necessity of cleaning the system up after each measurement.

- Contact: Annelie Heuser

5.5. SABR

Semantic-driven Analysis of BinaRies

KEYWORDS: Malware - Semantic - Binary analysis - Unsupervised graph clustering SCDG - Machine learning

FUNCTIONAL DESCRIPTION: Toolchain for binary analysis based on different techniques for capturing binaries' semantics and performing machine learning-assisted analysis. The primary use is malware analysis for malware detection and classification, either based on supervised and unsupervised learning.

This toolchain includes modules of the former BMA toolchain, specifically the SCDG extraction.

Our approach is based on artificial intelligence. We use concolic analysis to extract behavioral signatures from binaries in a form of system call dependency graphs (SCDGs). Our software can do both supervised and unsupervised learning. The former learns the distinctive features of different malware families on a large training set in order to classify the new binaries as malware or cleanware according to their behavioural signatures. In the unsupervised learning the binaries are clustered according to their graph similarity. The toolchain is orchestrated by an experiment manager that allows to easily setup, launch and view results of all modules of the toolchain.

- Contact: Olivier Zendra

5.6. ORQAL

ORQchestration of ALgorithms

KEYWORDS: Docker - Orchestration

FUNCTIONAL DESCRIPTION: ORQAL is a simple batch scheduler for docker cluster which can be used to remotely and without overhead in scientific experiment.

- Contact: Olivier Zendra

5.7. Side-channel deep learning evaluation platform

KEYWORDS: Deep learning - Evaluation

FUNCTIONAL DESCRIPTION: Our platform is based on several software. The first software permits to train a deep neural network and evaluate it for side-channel analysis, we evaluate our neural network with guessing entropy metrics. The second software is used for programming and communicating with the target devices, but we also develop a software to communicate with the equipment and made some measurement for side-channel analysis. The last software is used to make some attack and analysis of side-channel (e.g. made Correlation Power Analysis)

- Contact: Annelie Heuser

5.8. E-PAC

Evolving-PAcker Classifier

KEYWORDS: Packer classification - Incremental learning - Clustering - Malware - Obfuscation

FUNCTIONAL DESCRIPTION: E-PAC is an Evolving packer classifier that identifies the class of the packer used in a batch of packed binaries given in input. The software has the ability to identify both known packer classes and new unseen packer classes. After each update, the evolving classifier self-updates itself with the predicted packer classes.

The software is based on a semi-supervised machine learning system composed of an offline phase and an online phase. In the offline phase, a set of features is extracted from a collection of packed binaries provided with their ground truth labels, then a density-based clustering algorithm (DBSCAN) is used to group similar packers together with respect to a distance measure. In this step, the similarity threshold is tuned in order to form the clusters that fit the best with the the set of labels provided.

In the online phase, the software reproduces the same operations of features extraction and distances calculation with the incoming packed samples, then uses a customized version of the incremental clustering algorithm DBSCAN in order to classify them, either in knowns packer classes or fom new packer classes, or provisoirely leave them unclassified (notion of noise with DBSCAN).

The clusters formed after each update serve as a baseline for the application to self-evolve.

- Contact: Lamine Noureddine

6. New Results

6.1. Results for Axis 1: Vulnerability analysis

6.1.1. New Advances on Side-channel Distinguishers

Participants: Christophe Genevey Metat, Annelie Heuser.

A Systematic Evaluation of Profiling Through Focused Feature Selection.
Profiled side-channel attacks consist of several steps one needs to take. An important, but sometimes ignored, step is a selection of the points of interest (features) within side-channel measurement traces. A large majority of the related works start the analyses with an assumption that the features are preselected. Contrary to this assumption, here, we concentrate on the feature selection step. We investigate how advanced feature selection techniques stemming from the machine learning domain can be used to improve the attack efficiency. To this end, we provide a systematic evaluation of the methods of interest. The experiments are performed on several real-world data sets containing software and hardware implementations of AES, including the random delay countermeasure. Our results show that wrapper and hybrid feature selection methods perform extremely well over a wide range of test scenarios and a number of features selected. We emphasize L1 regularization (wrapper approach) and linear support vector machine (SVM) with recursive feature elimination used after chi-square filter (Hybrid approach) that performs well in both accuracy and guessing entropy. Finally, we show that the use of appropriate feature selection techniques is more important for an attack on the high-noise data sets, including those with countermeasures, than on the low-noise ones.

[3] Make Some Noise. Unleashing the Power of Convolutional Neural Networks for Profiled Side-channel Analysis. *Profiled side-channel analysis based on deep learning, and more precisely Convolutional Neural Networks, is a paradigm showing significant potential. The results, although scarce for now, suggest that such techniques are even able to break cryptographic implementations protected with countermeasures. In this paper, we start by proposing a new Convolutional Neural Network instance able to reach high performance for a number of considered datasets. We compare*

our neural network with the one designed for a particular dataset with masking countermeasure and we show that both are good designs but also that neither can be considered as a superior to the other one. Next, we address how the addition of artificial noise to the input signal can be actually beneficial to the performance of the neural network. Such noise addition is equivalent to the regularization term in the objective function. By using this technique, we are able to reduce the number of measurements needed to reveal the secret key by orders of magnitude for both neural networks. Our new convolutional neural network instance with added noise is able to break the implementation protected with the random delay countermeasure by using only 3 traces in the attack phase. To further strengthen our experimental results, we investigate the performance with a varying number of training samples, noise levels, and epochs. Our findings show that adding noise is beneficial throughout all training set sizes and epochs.

The Curse of Class Imbalance and Conflicting Metrics with Machine Learning for Side-channel Evaluations.

We concentrate on machine learning techniques used for profiled sidechannel analysis in the presence of imbalanced data. Such scenarios are realistic and often occurring, for instance in the Hamming weight or Hamming distance leakage models. In order to deal with the imbalanced data, we use various balancing techniques and we show that most of them help in mounting successful attacks when the data is highly imbalanced. Especially, the results with the SMOTE technique are encouraging, since we observe some scenarios where it reduces the number of necessary measurements more than 8 times. Next, we provide extensive results on comparison of machine learning and side-channel metrics, where we show that machine learning metrics (and especially accuracy as the most often used one) can be extremely deceptive. This finding opens a need to revisit the previous works and their results in order to properly assess the performance of machine learning in side-channel analysis.

[5] CC Meets FIPS: A Hybrid Test Methodology for First Order Side Channel Analysis.

Common Criteria (CC) and FIPS 140-3 are two popular side channel testing methodologies. Test Vector Leakage Assessment Methodology (TVLA), a potential candidate for FIPS, can detect the presence of side-channel information in leakage measurements. However, TVLA results cannot be used to quantify side-channel vulnerability and it is an open problem to derive its relationship with side channel attack success rate (SR), i.e., a common metric for CC. In this paper, we extend the TVLA testing beyond its current scope. Precisely, we derive a concrete relationship between TVLA and signal to noise ratio (SNR). The linking of the two metrics allows direct computation of success rate (SR) from TVLA for given choice of intermediate variable and leakage model and thus unify these popular side channel detection and evaluation metrics. An end-to-end methodology is proposed, which can be easily automated, to derive attack SR starting from TVLA testing. The methodology works under both univariate and multivariate setting and is capable of quantifying any first order leakage. Detailed experiments have been provided using both simulated traces and real traces on SAKURA-GW platform. Additionally, the proposed methodology is benchmarked against previously published attacks on DPA contest v4.0 traces, followed by extension to jitter based countermeasure. The result shows that the proposed methodology provides a quick estimate of SR without performing actual attacks, thus bridging the gap between CC and FIPS.

[13] Combining sources of side-channel information.

A few papers relate that multi-channel consideration can be beneficial for side-channel analysis. However, all were conducted using classical attack techniques. In this work, we propose to explore a multi-channel approach thanks to machine/deep learning. We investigate two kinds of multi-channel combinations. Unlike previous works, we investigate the combination of EM emissions from different locations capturing data-dependent leakage information on the device. Additionally, we consider the combination of the classical leaking signals and a measure of mostly the ambient noise. The knowledge of the ambient noise (due to WiFi, GSM, ...) may help to remove it from a noisy trace. To investigate these multi-channel approaches, we describe one option of how to extend a CNN architecture which takes as input multiple channels. Our results show that multi-channel networks *are suitable for side-channel analysis. However, if one channel alone already contains enough exploitable information to reach high effectiveness, naturally, the multi-channel approach cannot improve the performance further.*

6.1.2. Side-channel analysis on post-quantum cryptography

Participants: Tania Richmond, Yulliwas Ameur, Agathe Cheriere, Annelie Heuser.

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, in particular code-based cryptosystems.

During this year, we have set up a first side-channel experiment platform suited for embedded devices running code-based cryptosystems. Using this platform we exploited vulnerabilities of the syndrome computation present in some code-based algorithms.

6.1.3. Verification of IKEv2 protocol

Participants: Tristan Ninet, 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. IKEv2 guarantees security properties like mutual-authentication and secrecy of exchanged key. To obtain an IKEv2 implementation as secure as possible, we use model checking to verify the properties on the protocol specification, and software formal verification tools to detect implementation flaws like buffer overflows or memory leaks.

In previous analyses, IKEv2 has been shown to possess two authentication vulnerabilities that were considered not exploitable. We analyze the protocol specification using the Spin model checker, and prove that in fact the first vulnerability does not exist. In addition, we show that the second vulnerability is exploitable by designing and implementing a novel slow Denial-of-Service attack, which we name the Deviation Attack.

We propose an expression of the time at which Denial-of-Service happens, and validate it through experiment on the strongSwan implementation of IKEv2. As a counter-measure, we propose a modification of IKEv2, and use model checking to prove that the modified version is secure.

For ethical reasons we informed our country's national security agency (ANSSI) about the existence of the Deviation Attack. The security agency gave us some technical feedback as well as its approval for publishing the attack.

We then tackle formal verification directly applied to an IKEv2 source code. We already tried to analyze strongSwan using the Angr tool. However we found that the Angr was not mature yet for a program like strongSwan. We thus try other software formal verification tools and apply them to smaller and simpler source code than strongSwan: we analyze OpenSSL asn1parse using the CBMC tool and light-weight IP using the Infer tool. We find that CBMC does not scale to a large source code and that Infer does not verify the properties we want.

We explored more in-depth a formal technique and work towards the goal of verifying generic properties (absence of implementation flaws) on softwares like strongSwan.

Publications:

- [10] Model Checking the IKEv2 Protocol Using Spin
- [11] The Deviation Attack: A Novel Denial-of-Service Attack Against IKEv2

6.1.4. Software obfuscation

Participants: Alexandre Gonzalvez, Olivier Decourbe.

The limits of software obfuscation are not clear in practice. A protection based on opaque predicates can not be compatible with the control flow integrity property at low-level, due to the presence of indirect jumps in the instruction set architecture semantics. We propose a restricted instruction set architecture to overcome this limit. We argue for the adoption of restricted instruction set architecture for security-related computation. Publication:

- [9] A case against indirect jumps for secure programs

6.2. 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:

6.2.1. Malware Classification and clustering

Participants: Cassius Puodzius, Stefano Sebastio, Olivier Decourbe, Annelie Heuser, Olivier Zendra.

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.

Moreover, we explored the impact on the malware classification of several heuristics adoptable in the SCDGs building process and graph exploration. In particular, our purpose was to:

- identify quality characteristics and evaluation metrics of binary signatures based on SCDGs (and consequently the key properties of the execution traces), that characterize signatures able to provide high-precision malware classification
- optimize the performance of the SMT solver by designing a meta-heuristic able to select the best heuristic to tackle a specific sub-class of problem, study the impact of the configuration of the SMT solver and symbolic execution framework, and understand their interdependencies with the aim of efficiently extracting SCDGs in accordance with the identified quality metrics.

By adopting a Design of Experiments approach constituted by a full factorial experiment design and an Analysis of Variance (ANOVA) we have been able to pinpoint that, considering the graph metrics and their impact on the F-score, the litmus test for the quality of an SCDG-based classifier is represented by the presence of connected components. This could be explained considering how the graph mining algorithm (gSpan) works and the adopted similarity metric based on the number of common edges between the extracted signatures and the SCDG of the sample to classify. The results of the factorial experiments show that in our context tuning the symbolic execution is a very complex problem and that the sparsity of effect principle (stating that the system is dominated by the effect of the main factors and low-order-factor interactions) does not hold. The evaluation proved that the SMT solver is the most influential positive factor also showing an ability in reducing the impact of heuristics that may need to be enabled due to resource constraints (e.g., the max number of active paths). Results suggest that the most important factors are the disjoint union (as trace combination heuristic), and the our SMT optimization (through meta-heuristics) whereas other heuristics (such as min trace size and step timeout) have less impact on the quality of the constructed SCDGs.

During this year we build a end-to-end functional toolchain for supervised learning.

Furthermore, we have extended our approach to malware classification using unsupervised clustering. Preliminary results show that we are able to classify malware according to their behavioral properties without the need of any predefined labels.

6.2.2. Packers analysis

Participants: Lamine Nourredine, Cassius Puodzius, Stefano Sebastio, Annelie Heuser, Olivier Zendra.

Packing is a widespread tool to prevent static malware detection and analysis. Detecting and classifying the packer used by a given malware sample is fundamental to being able to unpack and study the malware, whether manually or automatically. Existing works on packing detection and classification has focused on effectiveness, but does not consider the efficiency required to be part of a practical malware-analysis workflow. This work studies how to train packing detection and classification algorithms based on machine learning to be both highly effective and efficient. Initially, we create ground truths by labeling more than 280,000 samples with three different techniques. Then we perform feature selection considering the contribution and computation cost of features. Then we iterate over more than 1,500 combinations of features, scenarios, and algorithms to determine which algorithms are the most effective and efficient, finding that a reduction of 1-2% effectiveness can increase efficiency by 17-44 times. Then, we test how the best algorithms perform against malware collected after the training data to assess them against new packing techniques and versions, finding a large impact of the ground truth used on algorithm robustness. Finally, we perform an economic analysis and find simple algorithms with small feature sets to be more economical than complex algorithms with large feature sets based on uptime/training time ratio.

A limit of supervised learning is to not be able to recognize classes that were not present in the ground truth. In the work's case above, this means that packer families for which a classifier has not been trained will not be recognized. In this work, we use unsupervised learning techniques, more particularly clustering, in order to provide information about packed malware with previously unknown packing techniques. Here, we build our own dataset of packed binaries, since in the previous work, it has been shown that the construction of the ground truth was fundamental in determining the effectiveness of the packing classification process. Choosing the right clustering algorithm with the right distance metric, dealing with different scales of features units, while being effective, efficient and robust are also majors parts of the current work.

During this year we have developed a toolchain of effective clustering of packers, in particular taking into account the possibility of evolution in packers. For this we derived and implemented new feature extraction strategies combined with incremental clustering algorithms.

6.3. (Coordination of the) H2020 TeamPlay Project, and Expression of Security Properties

Participants: Olivier Zendra, Yoann Marquer, Céline Minh, Nicolas Kiss, Annelie Heuser, Tania Richmond.

6.3.1. Overview & results

This work is done in the context of the TeamPlay EU project.

As mobile applications, the Internet of Things, and cyber-physical systems become more prevalent, so there is an increasing focus on energy efficiency of multicore computing applications. At the same time, traditional performance issues remain equally important. Increasingly, software designs need to find the best performance within some energy budget, often while also respecting real-time or other constraints, which may include security, data locality or system criticality, and while simultaneously optimising the usage of the available hardware resources.

While parallel multicore/manycore hardware can, in principle, ameliorate energy problems, and heterogeneous systems can help to find a good balance between execution time and energy usage, at present there are no effective analyses beyond user-guided simulations that can reliably predict energy usage for parallel systems, whether alone or in combination with timing information and security properties. In order to create energy-, time- and security- (ETS) efficient parallel software, programmers need to be actively engaged in decisions about energy usage, execution time and security properties rather than passively informed about their effects. This extends to design-time as well as to implementation-time and run-time.

In order to address this fundamental challenge, TeamPlay takes a radically new approach: by exploiting new and emerging ideas that allow non-functional properties to be deeply embedded within their programs, programmers can be empowered to directly treat energy ETS properties as first-class citizens in their parallel software. The concrete objectives of the TeamPlay project are:

1. To develop new mechanisms, along with their theoretical and practical underpinnings, that support direct language-level reasoning about energy usage, timing behaviour, security, etc.
2. To develop system-level coordination mechanisms that facilitate optimised resource usage for multicore hardware, combining system-level resource utilisation control during software development with efficient spatial and temporal scheduling at run-time.
3. To determine the fundamental inter-relationships between time, energy, security, etc. optimisations, to establish which optimisation approaches are most effective for which criteria, and to consequently develop multiobjective optimising compilers that can balance energy consumption against timing and other constraints.
4. To develop energy models for heterogeneous multicore architectures that are sufficiently accurate to enable high-level reasoning and optimisation during system development and at run-time.
5. To develop static and dynamic analyses that are capable of determining accurate time, energy usage and security information for code fragments in a way that can inform high-level programs, so achieving energy, time and security transparency at the source code level.
6. To integrate these models, analyses and tools into an analysis-based toolbox that is capable of reflecting accurate static and dynamic information on execution time and energy consumption to the programmer and that is capable of optimising time, energy, security and other required metrics at the whole system level.

7. To identify industrially-relevant metrics and requirements and to evaluate the effectiveness and potential of our research using these metrics and requirements.
8. To promote the adoption of advanced energy-, time- and security-aware software engineering techniques and tools among the relevant stake-holders.

Inria will exploit the results of the TeamPlay project in two main domains. First, they will strengthen and extend the research Inria has been carrying on low power and energy for embedded systems, especially for memory and wireless sensors networks. Second, they will complement in a very fitting way the research carried at Inria about security at a higher level (model checking, information theory).

The capability to express the energy and security properties at the developper level will be integrate in Inria own prototype tools, hence widening their applicability and the ease of experimentation. The use of energy properties wrt. evening of energy consumption to prevent information leakage, thus making side-channels attacks more difficult, is also a very promising path.

In addition, the methodological results pertaining to the development of embedded systems with a focus on low power and energy should also contribute to research lead at Inria in the domain of software engineering and advanced software engineering tools. Furthermore, security research lead at Inria will benefit from the security work undertaken by Inria and SIC in TeamPlay.

Overall, the project, with a strong industrial presence, will allow Inria to focus on matching concrete industrial requirements aiming at actual products, hence in providing more robust and validated results. In addition, the extra experience of working with industrial partners including SMEs will surely impact positively on Inria research methodology, making Inria research more attractive and influential, especially wrt. industry.

Finally, the results, both in terms of methodology and techniques, will also be integrated in the teaching Inria contributes to at Master level, in the areas of Embedded Systems and of Security.

The TeamPlay consortium agreement has been created by Inria, discussed with the various partners, and has been signed by all partners on 28 Feb. 2018. Inria has also distributed the partners initial share of the grant at the beginning of the project.

As WP7 (project management) leader and project coordinator, Inria was in charge of arranging general project meetings, including monthly meetings (tele-conferences), bi-annual physical meetings, boards meetings. During the first period, three exceptional physical meetings have been conducted, in addition to monthly project meetings: the kick-off meeting in Rennes from the 30th to the 31st of January 2018, the physical progress meeting has been conducted in Odense from the 26th to the 27th of June 2018, and the review in Brussels prepared the 19th of September 2018 and set the 17th of October 2018.

We have selected and set up utility tools for TeamPlay: shared notepads, mailing lists, shared calendars and collaborative repositories. We have ensured the timely production of the due deliverables. We set up the Project Advisory Board (PAB) with the aim of gathering external experts from both academia and industry, covering a wide range of domains addressed by TeamPlay. Finally, we ensured good working relationships (which can implicate conflict resolution when needed), monitored the overall progress of the project, and reported to the European Commission on technical matters and deliverables.

We also organized a tooling meeting in Hamburg in October the 30th, to discuss the relation between the tools from different partners, e.g. Idris from the University of St Andrews, the WCC compiler developed in the Hamburg University of Technology, or the coordination tool developed in the University of Amsterdam.

Measuring security, unlike measuring other more common non-functional properties like time or energy, is still very much in its infancy. For example, time is often measured in seconds (or divisions thereof), but security has no widely agreed, well-defined measurement. It is thus one goal of this project, especially for SIC and Inria, to design (necessarily novel) security measurements, and have them implemented as much as possible throughout the set of development tools.

Measuring security by only one value however seems impossible or may be meaningless. More precisely, if security could be defined overall by only one measurement, the latter would be a compound (i.e. an aggregation) of several more specialized measurement. Indeed, security encompasses many aspects of interest:

1. By allowing communications between different systems, security properties should be guaranteed in order to prevent low-level users from determining anything about high-level users activity, or in the case of public communication channels in a hostile environment, to evaluate vulnerability to intruders performing attacks on communications.
   1. *Confidentiality* (sometimes called *secrecy*) properties like non-interference (and many) variants can be described by using an information-flow policy (e.g. high- and low-level users) and studying traces of user inputs.
   2. *Vulnerability* captures how a system is sensible to attacks on communications (e.g. stealing or faking information on a public channel).
2. A *side-channel* is a way of transmitting informations (purposely or not) to another system out of the standard (intended) communication channels. *Side-channel attacks* rely on the relationship between information leaked through a side-channel and the secret data to obtain confidential (non-public) information.
   1. *Entropy* captures the uncertainty of the attacker about the secret key. The attacker must be able to extract information about the secret key through side-channel measurements, which is captured by the *attacker's remaining uncertainty* value, which can be computed by using heuristic techniques. The attacker must also be able to effectively recover the key from the extracted information, which is expressed by the *min-entropy leakage*, and refined by the *g-leakage* of a gain function.
   2. The power consumption of a cryptographic device can be analyzed to extract the secret key. This is done by using several techniques: visual examination of graphs of the current (*Simple Power Analysis*), by exploiting biases in varying power consumption (*Differential Power Analysis*), or by using the correlation coefficient between the power samples and hypotheses (*Correlation Power Analysis*).
   3. Usual security properties guarantee only the input-output behavior of a program, and not its execution time. Closing *leakage through timing* can be done by disallowing while-loops and if-commands to depend on high security data, or by padding the branches so that the external observer cannot determine which branch was taken.
   4. Finally, the correlation between the patterns of the victim's execution and the attacker's observations is formalized as a metric called the *Side-channel Vulnerability Factor*, which is refined by the *Cache Side-channel Vulnerability* for cache attacks.
3. A cryptographic scheme should be secure even if the attacker knows all details about the system, with the exception of the secret keys. In particular, the system should be secure when the attacker knows the encryption and decryption algorithms.
   1. In modern cryptography, the security level (or security strength) is given by the *work factor*, which is related to its key-length and the number of operations necessary to break a cryptographic scheme (try all possible combinations of the key). An algorithm is said to have a "security level of n bits" if the best known attack requires 2^n steps. This is a quite natural definition because symmetric algorithms with a security level of n have a key of length n bits.
   2. The relationship between cryptographic strength and security is not as straightforward in the asymmetric case. Moreover, for symmetric algorithms, a key-length of 128 bits provides an estimated long term security (i.e. several decades in the absence of quantum computer) regarding brute-force attacks. To reach an estimated long term security even with quantum computers, a key-length of 256 bits is mandatory.

Inria is implementing side-channel countermeasures (hiding) into the WCET-aware C Compiler (WCC) developed by the Hamburg University of Technology (TUHH). A research visit to TUHH was arranged with the aim at learning how to work on WCC (TUHH and WCC infrastructure, WCC developers best practices, etc.). Inria will use compiler-based techniques to prevent timing leakages and power leakages.

For instance, in a conditional branching `if` b `then` $P_1(x)$ `else` $P_2(x)$, measuring the execution time or the power profile may allow to know whether the branch P_1 or P_2 have been chosen to manipulate the value x, thus to obtain the secret value b. To prevent timing leakage, P_1 and/or P_2 can be padded (i.e. dummy instructions are added) in order to obtain the worst-case execution time in both branches.

But this does not prevent information leakage from power profile. A stronger technique, from a security point of view, could be to add a dummy variable y and duplicate the code such that $y = x;$ `if` b `then` $P_1(x); P_2(y)$ `else` $P_1(Y); P_2(x)$ always performs the operations of P_1 then the operations of P_2. But the execution time is now the sum and not the worst-case of both branches, thus trading execution time to increase security.

Finally, the initialization $y = x$ can be detected, and the previous solution is still vulnerable to fault injections. Some algorithms like the Montgomery Ladder are more protected against these attacks because both variables x and y are entangled during the execution. We hope to generalize this property to a wider set of algorithms, or to automatically detect the properties required from the original code in order to transform it into a “ladderized" version with higher security level.

6.3.2. Publication

- Type-Driven Verification of Non-functional Properties [8].

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

- CISCO (http://www.cisco.com) contract (2017–2019) to work on graph analysis of malware

7.2. 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

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. 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.
- ANR AHMA, Automated Hardware Malware Analysis, 3,5 years, JCJC.

8.1.2. DGA

- PhD grant for Nisrine Jafri (2016–2019),
- PhD grant for Lamine Noureddine (2017-2020)
- PhD grant for Christophe Genevey Metat (2018-2021)
- PhD grant for Cassius De Oliveira Puodzius (2019-2022)

8.2. European Initiatives

8.2.1. ENABLE-S3 (352)

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 Sl (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: Olivier Zendra

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.

Within ENABLE-S3, the contribution of the TAMIS team consists in in proposing a generic method to evaluate complex automotive-oriented systems for automation (perception, decision-making, etc.). The method is based on Statistical Model Checking (SMC), using specifically defined Key Performance Indicators (KPIs), as temporal properties depending on a set of identified metrics. By feeding the values of these metrics during a large number of simulations, and the properties representing the KPIs to our statistical model checker, we evaluate the probability to meet the KPIs. We applied this method to two different subsystems of an autonomous vehicles: a perception system (CMCDOT framework) and a decision-making system. We show that the methodology is suited to efficiently evaluate some critical properties of automotive systems, but also their limitations.

In 2019, in TAMIS, Olivier Zendra and Eduard Baranov were involved in this project. The project supported one postdoc in TAMIS starting in 2017.

8.2.2. TeamPlay (653)

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 (Denmark), 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

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 Olivier Zendra and Annelie Heuser.

8.2.3. SUCCESS

Title: SUCCESS: SecUre aCCESSibility for the internet of things

Program: CHIST-ERA 2015

Duration: 10/2016 - 10/2019

Coordinator: Middlesex University (UK)

Partners:

Middlesex University, School of Science and Technology (UK); Inria, TAMIS (France); Université Grenoble Alpes, Verimag (France); University of TWENTE, (Netherlands)

Inria contact: Ioana Cristescu

The objectives of the SUCCESS project 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 its 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 2019, in the TAMIS team, Delphine Beaulaton, Najah Ben Said, Ioana Cristescu and Olivier Zendra were involved in this project.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events: Selection

9.1.1.1. Member of Conference Steering Committees

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)

9.1.1.2. Member of the Conference Program Committees

Annelie Heuser was a member of the following conference program committees: CARDIS 2019, COSADE 2019, PROOFS 2019.
Olivier Zendra was a member of the program committee of ICOOOLPS 2019.

9.1.1.3. Reviewer

Tania Richmond was a reviewer for TCHES 2019 and TCHES 2020.
Alexandre Gonzalvez was a reviewer for TCHES 2019.
Matthieu Mastio was a reviewer for DATE2020.

9.1.2. Journal

9.1.2.1. Member of the Editorial Boards

Annelie Heuser was a member of TCHES 2019.

9.1.2.2. Reviewer - Reviewing Activities

Annelie Heuser was a reviewer of several papers in Journal of Cryptographic Engineering, Transactions on Information Forensics & Security, IEEE Transactions on Circuits and Systems, Transactions on Information Forensics & Security.
Tania Richmond was a reviewer for Transactions on Information Forensics and Security (TIFS)

9.1.3. Invited Talks

Annelie Heuser was invited to present at the seminar of Paris 8 on “Mathématiques Discrètes, Workshop Workshop on Randomness and Arithmetics for Cryptography on Hardware, Codes et Cryptographie”, GDR Sécurité - Sécurité, fiabilité et test des SoC2, the Summer School on real-world crypto and privacy.
Annelie Heuser was invited to the Dagstuhl seminar on Secure Composition for Hardware Systems.
Annelie Heuser was invited to the Lorentzcenter on AI+Sec: Artificial Intelligence and Security, and SHARD: Bridging the Gap Between Software and Hardware Security.

9.1.4. Expertise

Tania Richmond reviewed for the HiPEAC Collaboration Grants 2019.
Annelie Heuser reviewed an application to receive a national application grant (inside Europe).
Olivier Zendra is a CIR expert for the MENESR.
Olivier Zendra participated to the CRHC and CRCN national juries for Inria as a member of Inria’s evaluation committee.
Olivier Zendra is a member of the editorial board and co-author of the “HiPEAC Vision 2021”

9.1.5. Research Administration

Olivier Zendra is a member of Inria’s evaluation committee.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Christophe Genevey-Metat taught:

- Programming in Java to Licence 1, University Rennes 1,
- Introduction at Security to Master 1, University Rennes 1.

Cassius de Oliveira Puodzius taught:

- 24h of Reverse Engineering + Malware analysis for the Introduction at Security to Master1 at Rennes 1,
- 6h of Java Programming to Licence1 at ECAM Rennes.

Tania Richmond taught Physical Attacks in "Préparation à l’Agrégation" en Sciences de l’Ingénieur, ENS Rennes.

9.2.2. Supervision

- PhD in progress: Christophe Genevey Metat (Rennes 1): , October 2018, Jean-Marc Jezequel, Benoit Gerard, Annelie Heuser and Clementine Maurice
- PhD in progress : Alexandre Gonzalvez, On Obfuscation via crypto primitives, April 2016, Caroline Fontaine.
- PhD defended in 2019 : Nisrine Jafri (Rennes1), On fault Injection detection with MC of Binary code, Axel Legay and Jean-Louis Lanet.
- PhD in progress : Tristan Ninet (Rennes 1), Vérification formelle d'une implémentation de la pile protocolaire IKEv2, December 2016, Stéphanie Delaune, Romaric Maillard and Olivier Zendra
- PhD in progress: Lamine Nouredine (Rennes1); Automatic techniques for packing detection, classification and unpacking to stop malware propagation, November 2017, Stephane Ubeda, Olivier Zendra, Annelie Heuser.
- PhD in progress: Cassius de Oliveira Puodzius (Rennes1); Threat malware analysis, February 2019, Ludovic Me, Olivier Zendra, Annelie Heuser.
- PhD in progress: Duc Phuc Pham (Rennes1); Malware analysis through side-channel information, May 2019, Jean-Louis Lanet, Annelie Heuser, Olivier Zendra.

9.2.3. Juries

- Annelie Heuser was a referee for the PhD defense of Nisrine Jafri (University Rennes).
- Olivier Zendra was a member of the jury for the PhD defense of Delphine Beaulaton at University of Southern Brittany in Vannes.

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] F. DOLD.*The GNU Taler system : practical and provably secure electronic payments*, Université Rennes 1, February 2019, https://tel.archives-ouvertes.fr/tel-02138082

[2] N. JAFRI.*Formal fault injection vulnerability detection in binaries : a software process and hardware validation*, Université Rennes 1, March 2019, https://tel.archives-ouvertes.fr/tel-02385208

Articles in International Peer-Reviewed Journal

[3] J. KIM, S. PICEK, A. HEUSER, S. BHASIN, A. HANJALIC.*Make Some Noise Unleashing the Power of Convolutional Neural Networks for Profiled Side-channel Analysis*, in "IACR Transactions on Cryptographic Hardware and Embedded Systems", May 2019 [*DOI :* 10.13154/TCHES.V2019.I3.148-179], https://hal.inria.fr/hal-02010599

[4] S. PICEK, A. HEUSER, A. JOVIC, S. BHASIN, F. REGAZZONI.*The Curse of Class Imbalance and Conflicting Metrics with Machine Learning for Side-channel Evaluations*, in "IACR Transactions on Cryptographic Hardware and Embedded Systems", August 2019, vol. 2019, n^o 1, p. 1-29 [*DOI :* 10.13154/TCHES.V2019.I1.209-237], https://hal.inria.fr/hal-01935318

[5] D. B. ROY, S. BHASIN, S. GUILLEY, A. HEUSER, S. PATRANABIS, D. MUKHOPADHYAY.*CC Meets FIPS: A Hybrid Test Methodology for First Order Side Channel Analysis*, in "IEEE Transactions on Computers", March 2019, vol. 68, n^o 3, p. 347-361 [*DOI :* 10.1109/TC.2018.2875746], https://hal.inria.fr/hal-02413209

International Conferences with Proceedings

[6] M. BARBIER, A. RENZAGLIA, J. QUILBEUF, L. RUMMELHARD, A. PAIGWAR, C. LAUGIER, A. LEGAY, J. IBAÑEZ-GUZMÁN, O. SIMONIN.*Validation of Perception and Decision-Making Systems for Autonomous Driving via Statistical Model Checking*, in "IV 2019 - 30th IEEE Intelligent Vehicles Symposium", Paris, France, IEEE, June 2019, p. 252-259 [*DOI :* 10.1109/IVS.2019.8813793], https://hal.inria.fr/hal-02127889

[7] P. BOUTILLIER, I. CRISTESCU, J. FERET.*Counters in Kappa: Semantics, Simulation, and Static Analysis*, in "ESOP 2019 - 28th European Symposium on Programming", Prague, Czech Republic, Springer, April 2019, p. 176-204 [*DOI :* 10.1007/978-3-030-17184-1_7], https://hal.inria.fr/hal-02397876

[8] C. BROWN, A. D. BARWELL, Y. MARQUER, C. MINH, O. ZENDRA.*Type-Driven Verification of Non-functional Properties*, in "PPDP 2019 - 21st International Symposium on Principles and Practice of Declarative Programming", Porto, Portugal, ACM Press, October 2019, p. 1-15 [*DOI :* 10.1145/3354166.3354171], https://hal.inria.fr/hal-02314723

[9] A. GONZALVEZ, R. LASHERMES.*A case against indirect jumps for secure programs*, in "SSPREW-9 2019 - 9th Software Security, Protection, and Reverse Engineering Workshop", San Juan, United States, SSPREW9 '19: Proceedings of the 9th Workshop on Software Security, Protection, and Reverse Engineering, December 2019, p. 1-10, https://hal.archives-ouvertes.fr/hal-02382711

[10] T. NINET, A. LEGAY, R. MAILLARD, L.-M. TRAONOUEZ, O. ZENDRA.*Model Checking the IKEv2 Protocol Using Spin*, in "PST 2019 - 17th International Conference on Privacy, Security and Trust", Fredericton, Canada, August 2019, p. 1-9, https://hal.inria.fr/hal-02062292

[11] T. NINET, A. LEGAY, R. MAILLARD, L.-M. TRAONOUEZ, O. ZENDRA.*The Deviation Attack: A Novel Denial-of-Service Attack Against IKEv2*, in "TrustCom 2019 - 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications", Rotorua, New Zealand, IEEE, August 2019, p. 1-8, https://hal.inria.fr/hal-01980276

Conferences without Proceedings

[12] F. DÉCHELLE, B. LEBON, O. ZENDRA.*MASSE: Modular Automated Syntactic Signature Extraction*, in "RESSI 2019 - Rendez-vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information", Erquy, France, May 2019, 1, https://hal.inria.fr/hal-02159947

[13] C. GENEVEY-METAT, B. GÉRARD, A. HEUSER.*Combining sources of side-channel information*, in "C&ESAR 2019", Rennes, France, November 2019, https://hal.archives-ouvertes.fr/hal-02456646

Books or Proceedings Editing

[14] M. DURANTON, K. DE BOSSCHERE, B. COPPENS, C. GAMRAT, M. GRAY, H. MUNK, E. OZER, T. VARDANEGA, O. ZENDRA (editors). *The HiPEAC Vision 2019*, HiPEAC CSA, January 2019, 178, https://hal.inria.fr/hal-02314184

Other Publications

[15] C. AUBERT, I. CRISTESCU.*History-Preserving Bisimulations on Reversible Calculus of Communicating Systems*, February 2019, https://arxiv.org/abs/1804.10355 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01778656

[16] S. PICEK, A. HEUSER, C. ALIPPI, F. REGAZZONI.*When Theory Meets Practice: A Framework for Robust Profiled Side-channel Analysis*, February 2019, working paper or preprint, https://hal.inria.fr/hal-02010603

[17] T. RICHMOND, A. HEUSER, B. GÉRARD.*Side-Channel Analysis of Post-Quantum Cryptography*, January 2019, 1, SecDays 2019 - Security Days, Poster, https://hal.inria.fr/hal-02018859

References in notes

[18] 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

Project-Team TEA

Time, Events and Architectures

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Embedded and Real-time Systems

Table of contents

Project-Team TEA

Creation of the Team: 2014 January 01, updated into Project-Team: 2015 January 01

Keywords:

Computer Science and Digital Science:

A1.2.5. - Internet of things
A1.2.7. - Cyber-physical systems
A1.5.2. - Communicating systems
A2.1.1. - Semantics of programming languages
A2.1.4. - Functional programming
A2.1.6. - Concurrent programming
A2.1.9. - 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. - Formal method for 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. - Calculability and computability
A8.1. - Discrete mathematics, combinatorics
A8.3. - Geometry, Topology

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. Team, Visitors, External Collaborators

Research Scientists

Jean-Pierre Talpin [Team leader, Inria, Senior Researcher, HDR]
Thierry Gautier [Inria, Researcher]
Rajesh Kumar Gupta [University of California San Diego, Inria Chair]
Shuvra Bhattacharyya [University of Maryland, Insa-Inria Chair]

Technical Staff

Loic Besnard [CNRS SED, Senior Engineer]
Armelle Mozziconacci [CNRS, Project Assistant]

PhD Students

Lucas Franceschino [Inria]
Stephane Kastenbaum [Mitsubishi Electric, from Oct 2019]
Simon Lunel [Inria, until Jan 2019]
Jean Joseph Marty [Inria]
Liangcong Zhang [China Scolarship Council]

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: modeling and programming formalisms that represent software, hardware and physics significantly alter this perception of time. Therefore, 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 functionalities, thanks 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, such as for instance Simulink and Matlab. 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. CPS design is, to date, mostly executed in this ad-hoc manner, without sound, mathematically grounded, integrative methodology. A new science of CPS design will allow to create machines with complex dynamics and high control reliability, and apply to new industries and applications, such as IoT or edge devices, 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 [0][0][0], 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

[0] *Matlab/Simulink*, https://fr.mathworks.com/products/simulink.html
[0] *Ptolemy*, http://ptolemy.eecs.berkeley.edu
[0] *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.

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 is inherently capable of describing multi-clock abstractions of GALS systems. Polychrony is implemented in the data-flow specification language Signal, available in the Eclipse project POP [0] and in the CCSL standard [0] available from the TimeSquare project. Both 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.

[0] *Polychrony on Polarsys*, https://www.polarsys.org/projects/polarsys.pop

[0] *Clock Constraints in UML/MARTE CCSL*. C. André, F. Mallet. RR-6540. Inria, 2008. http://hal.inria.fr/inria-00280941

3.2. Timed Modeling

To formalize timed semantics for system design, we shall rely on algebraic representations of time as clocks found in previous works and introduce a paradigm of "time system" (types that 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 timed concurrency.

In particular, the principle of refinement type systems [0], 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 [0]. 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 requires the use of assume-guarantee reasoning to allow modularity and to facilitate composition by behavioral sub-typing, in the spirit of the (static) contract-based formalism proposed by Passerone et al. [0]. Verification problems encompassing heterogeneously timed specifications are common and of great variety: checking correctness between abstract (e.g. the synchronous hypothesis) and concrete time models (e.g. real-time architectures) relates to desynchronisation (from synchrony to asynchrony) and scheduling analysis (from synchronous data-flow 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 to use abstraction and refinement mechanisms for time models (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 problems such as these integrating constraints of battery capacity, on-board CPU performance, available memory resources, software schedulability, to logical software correctness and plant controllability.

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 [0], SysML [0], UML MARTE [0].

In system design, an architectural specification serves several important purposes. First, it breaks down a system model into components of manageable size and complexity, 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. Connections between components, which specify how components interact with each other, help propagate the effects of a change in one component to the linked components.

[0] *Abstract Refinement Types*. N. Vazou, P. Rondon, and R. Jhala. European Symposium on Programming. Springer, 2013.
[0] *LTL types FRP*. A. Jeffrey. Programming Languages meets Program Verification.
[0] *A contract-based formalism for the specification of heterogeneous systems*. L. Benvenistu, et al. FDL, 2008
[0] *Architecture Analysis and Design Language*, AS-5506. SAE, 2004. http://standards.sae.org/as5506b
[0] *System modeling Language*. OMG, 2007. http://www.omg.org/spec/SysML
[0] *UML Profile for MARTE*. OMG, 2009. http://www.omg.org/spec/MARTE

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 domain-specific languages such as Prelude [0] 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 [0].

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 [0] 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 [0]. 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 [0][0].

Abstract scheduling, or the use of abstraction and refinement techniques in scheduling borrowed to the theory of abstract interpretation [0] 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 focus.

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.

3.5. Verified programming for system design

The IoT is a network of devices that sense, actuate and change our immediate environment. Against this fundamental role of sensing and actuation, design of edge devices often considers actions and event timings to

[0] *The Prelude language*. LIFL and ONERA, 2012. http://www.lifl.fr/~forget/prelude.html
[0] *PyECDAR, timed games for timed specifications*. Inria, 2013. https://project.inria.fr/pyecdar
[0] *A survey of hard real-time scheduling for multiprocessor systems*. R. I. Davis and A. Burns. *ACM Computing Survey* 43(4), 2011.
[0] *Buffer minimization in EDF scheduling of data-flow graphs*. A. Bouakaz and J.-P. Talpin. LCTES, ACM, 2013.
[0] *ADFG for the synthesis of hard real-time applications*. A. Bouakaz, J.-P. Talpin, J. Vitek. ACSD, IEEE, June 2012.
[0] *Design of SCJ Level 1 Applications Using Affine Abstract Clocks*. A. Bouakaz and J.-P. Talpin. SCOPES, ACM, 2013.
[0] *La vérification de programmes par interprétation abstraite*. P. Cousot. Séminaire au Collège de France, 2008.

be primarily software implementation issues: programming models for IoT abstract even the most rudimentary information regarding timing, sensing and the effects of actuation. As a result, applications programming interfaces (API) for IoT allow wiring systems fast without any meaningful assertions about correctness, reliability or resilience.

We make the case that the "API glue" must give way to a logical interface expressed using contracts or refinement types. Interfaces can be governed by a calculus – a refinement type calculus – to enable reasoning on time, sensing and actuation, in a way that provides both deep specification refinement, for mechanized verification of requirements, and multi-layered abstraction, to support compositionality and scalability, from one end of the system to the other.

Our project seeks to elevate the “function as type” paradigm to that of “system as type”: to define a refinement type calculus based on concepts of contracts for reasoning on networked devices and integrate them as cyber-physical systems [0]. An invited paper [0] outlines our progress with respect to this aim and plans towards building a verified programming environment for networked IoT devices: we propose a type-driven approach to verifying and building safe and secure IoT applications.

Accounting for such constrains in a more principled fashion demands reasoning about the composition of all the software and hardware components of the application. Our proposed framework takes a step in this direction by (1) using refinement types to make make physical constraints explicit and (2) imposing an event-driven programming discipline to simplify the reasoning of system-wide properties to that of an event queue. In taking this approach, our approach would make it possible for a developer to build a verified IoT application by ensuring that a well-typed program cannot violate the physical constraints of its architecture and environment.

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 focuses 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.

[0] Refinement types for system design. Jean-Pierre Talpin. FDL’18 keynote.

[0] Steps toward verified programming of embedded computing systems. Jean-Pierre Talpin, Jean-Joseph Marty, Deian Stefan, Shravan Nagarayan, Rajesh Gupta, DATE’18.

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 the data-flow language Signal 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 for virtual prototyping of 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 electro-mechanical 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 the software or 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

Loïc Besnard was promoted to the rank of Senior Engineer Exceptional Class by CNRS, acknowledging his remarkable career of research engineer as principal developer of Signal and Polychrony, as project manager and integrator with project teams EPATR (Signal), ESPRESSO (Polychrony), TEA (ADFG) and PACAP (Heptane).

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 model-driven 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 open-source 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[0]: 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

0

SIGNAL V4-Inria version: Reference Manual. Besnard, L., Gautier, T. and Le Guernic, P. http://www.irisa.fr/espresso/Polychrony, 2010 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, Jean-Pierre Talpin, Shuvra Bhattacharyya, Alexandre Honorat, Hai Nam Tran.

ADFG (Affine DataFlow Graph) synthesizes scheduling parameters for real-time systems modeled as synchronous data flow (SDF), cyclo-static dataflow (CSDF), and ultimately cyclo-static dataflow (UCSDF) graphs. It aims at mitigating the trade-off between throughput maximization and total buffer size minimization. The synthesizer inputs are a 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 synthesized scheduling parameters such as tasks periods, offsets, processor bindings, priorities, buffer initial markings and buffer sizes. ADFG was originally implemented by Adnan Bouakaz [0]. It is now being collaboratively developed with team Tea, Hai Nam Tran (UBO) Alexandre Honorat (INSA) and Shuvra Bhattacharyya (UMD/INSA/Inria).

ADFG is extended to support automated code generation of the computed buffer sizes and scheduling parameters for dataflow applications that are implemented in the Lightweight Dataflow Environment (LIDE) [0]. LIDE is a flexible, lightweight design environment that allows designers to experiment with dataflow-based implementations directly. LIDE actors and buffers (FIFOs) can be initialized with parameters, including buffer sizes. The usage of LIDE allows a systematic way to instantiate dataflow graphs with the buffer size parameters computed by ADFG.

[0] Real-Time Scheduling of Dataflow Graphs. A. Bouakaz. Ph.D. Thesis, University of Rennes 1, 2013.

[0] S. Lin, Y. Liu, K. Lee, L. Li, W. Plishker, and S. S. Bhattacharyya. 2017. The DSPCAD framework for modeling and synthesis of signal processing systems. Handbook of Hardware/Software Codesign (2017), 1185–1219.

Actor models and scheduling algorithms in ADFG have been extended to investigate the contention-aware scheduling problem on multi/many-core architectures. The problem we tackled is that the scheduler synthesis for these platforms must account for the non-negligible delay due to shared memory accesses. We exploited the deterministic communications exposed in SDF graphs to account for the contention and further optimize the synthesized schedule. Two solutions are proposed and implemented in ADFG: contention-aware and contention-free scheduling synthesis. In other words, we either take into account the contention and synthesize a contention-aware schedule or find a one that results in no contention.

ADFG is extended to apply a transformation known as partial expansion graphs (PEG). This transformation can be applied as a pre-processing stage to improve the exploitation of data parallelism in SDF graphs on parallel platforms. In contrast to the classical approaches of transforming SDF graphs into equivalent homogeneous forms, which could lead to an exponential increase in the number of actors and excessive communication overhead, PEG-based approaches allow the designer to control the degree to which each actor is expanded. A PEG algorithm that employs cyclo-static data flow techniques is developed in ADFG. Compared to existing PEG-based approach, our solution requires neither buffer managers nor split-join actors to coordinate data production and consumption rates. This allows us to reduce the number of added actors and communication overhead in the expanded graphs.

7.2. Parallel Composition and Modular Verification of Computer Controlled Systems in Differential Dynamic Logic

Participants: Jean-Pierre Talpin, Benoit Boyer, David Mentre, Simon Lunel, Stefan Mitsch.

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 the decomposition of systems into components enhanced with compositional contract reasoning.

The work of A. Platzer on Differential Dynamic Logic ($d\mathcal{L}$) held our attention [0]. This formalism is built upon the Dynamic Logic of V. Pratt and augmented with the possibility of expressing Ordinary Differential Equations (ODEs). Combined with the ability of Dynamic Logic to specify and verify hybrid programs, $d\mathcal{L}$ is particularly adapted to model cyber-physical systems. The proof system associated with the logic is implemented into the theorem prover KeYmaera X. Aimed toward automation, it is a promising tool to spread formal methods in industry.

Computer-Controlled Systems (CCS) are a subclass of hybrid systems where the periodic relation of control components to time is of paramount importance. Since they additionally are at the heart of many safety-critical devices, it is of primary importance to correctly model such systems and to ensure they function correctly according to safety requirements. Differential dynamic logic $d\mathcal{L}$ is a powerful logic to model hybrid systems and to prove their correctness. We contributed a compositional modeling and reasoning framework to $d\mathcal{L}$ that separates models into components with timing guarantees, such as reactivity of controllers and controllability of continuous dynamics. Components operate in parallel, with coarse-grained interleaving, periodic execution and communication. We present techniques to automate system safety proofs from isolated, modular, and possibly mechanized proofs of component properties parameterized with timing characteristics.

7.3. Multithreaded code generation for process networks

Participants: Loïc Besnard, Thierry Gautier.

As part of an in-depth comparison of process models, we have recently revisited the relation between the model of asynchronous dataflow represented by Kahn Process Networks (KPNs) and that of synchronous dataflow represented by the polychronous model of computation. In particular, we have precisely described in which conditions polychronous programs can be seen as KPNs. In this context, we have considered different cases

[0] *Differential Dynamic Logic for Hybrid Systems*, André Platzer, http://symbolaris.com/logic/dL.html

of process networks, including so-called "polyendochronous processes". Under some conditions expressed by clock equation systems, (networks of) processes exhibiting polyhierarchies of clocks are polyendochronous and, as compositions of endochronous processes, may be seen as KPNs.

Based on this characterization, we have developed in the open-source Polychrony toolset a new strategy of code generation for such (polyendochronous) process networks. Typically, after the clock calculus, a program P is organized as a composition of processes, $P = (| \ P1 \ | \ P2 \ | \ ... \ | \ Pn \ |)$, each one structured around a clock tree. When P is characterized as polyendochronous, it contains generally clock constraints such as $Clk1 = Clk2$, with $Clk1$ being a clock in the subtree corresponding to $P1$ and $Clk2$ a clock in the subtree corresponding to $P2$.

Such a constraint induces a synchronization between two parts ($P1$, $P2$) of the program when $Clk1$ or $Clk2$ occurs. The principle of the code generation for polyendochronous processes is based on the existing distributed code generation, but with the additional resynchronization of parts of the application induced by the constraints on clocks ($Clk1$, $Clk2$) not placed in the same clock trees. For distributed code generation, it is considered that each (clock) hierarchy will run on a specific processor. In this case, the purpose is mainly to partition the application, and the processors will be virtual ones.

The code generation of each partition consists in the definition of several tasks: one task per cluster (a cluster being a subpart that may be executed as soon as its inputs are available, without any communication with the external world); one task per input/output of the partition; one task for the cluster of state variables; one task that manages the steps. Synchronization between these tasks is obtained by semaphores (one semaphore per task). This code generation technique for the class of networks called "polyendochronous processes" has been added in the Polychrony toolset (http://polychrony.inria.fr) and a paper describing the comparison of process models is currently in submission.

7.4. Type theory for modular static analysis of system programs

Participants: Lucas Franceschino, Jean-Pierre Talpin, David Pichardie.

This Ph.D. project is about formal verification, with system programming applications in mind. Formal methods are essential for safety-critical software (i.e. transport and aeronautic industry). In the same time, more and more programming languages with a strong type system arise (such as Haskell, Rust, ML, Coq, F*, Idris...).

Formal methods come in different flavors: type theory, abstract interpretation, refinement types. Each of these "flavors" are both theoretical fields and are also being implemented concretely: *Astrée* ou *Verasco* for abstract interpretation, *Coq*, *Agda*, *F** or *Idris* dependent types, and *Liquid Haskell* for refinement types.

Our approach consists in positioning ourselves between type theory and abstract interpretation, and to leverage the power of both. The main intuition behind this idea is that abstract interpretation, suffering from expressiveness, would bring *invariant inference* power, while strong type systems, requiring manual annotations and proofs, would bring *expressivity*.

We formalized how one can enrich a weakest precondition calculus (WP) with an abstract interpreter. This work takes the shape of a WP calculus transformer: given a WP calculus, we generically construct a brand new WP calculus that produces easier (but sound, still) weakest preconditions, thanks to abstract interpretation.

Concretely, our work is being implemented as an F* effect transformer that leverage Verasco capabilities, for a low-level subset of F*, namely Low*.

7.5. Verified information flow of embedded programs

Participants: Jean-Joseph Marty, Lucas Franceschino, Niki Vazou, Jean-Pierre Talpin.

This PhD project is about applying refinement types theory to verified programming of applications and modules of library operating systems, such as unikernels, for embedded devices of the Internet of Things (IoT): TinyOS, Riot, etc. Our topic has focused on developing a model of information flow control using labeled input-outputs (LIO) implemented using F☆: project Lio☆.

As part of the development of Lio☆, we implemented a library that, thanks to static verification, ensures the containment of information in relation to a parameterized policy for information flow control. In collaboration with Niki Vazou (IMDEA) and Lucas Franceschino we have formalized and developed an automatic method to prove non-interference in Meta☆. Using the Kremlin code generator, programs using Lio☆ can be compiled into C code and run natively on embedded low-resource-constrained devices, without the need for additional runtime system.

In parallel we continued our collaboration with the ProgSys team on a second, now discontinued, project: Gluco☆. The goal of this project was to evaluate the capabilities to use the F* programming language to program an entire system by taking into account its software, hardware and physical constraints using type refinements [0].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Inria – Mitsubishi Electric framework program (2018+)

Title: Inria – Mitsubishi Electric framework program

Inria principal investigator: Jean-Pierre Talpin

International Partner: Mitsubishi Electric R&D Europe (MERCE)

Duration: 2018+

Abstract: Following up the fruitful collaboration of TEA with the formal methods group at MERCE, Inria and Mitsubishi Electric signed a center-wide collaboration agreement, which currently hosts projects with project-teams Sumo and Tea, as well as Tocata.

8.1.2. Mitsubishi Electric R&D Europe (2019-2022)

Title: A logical framework to verify requirements of hybrid system models

Inria principal investigator: Jean-Pierre Talpin, Stéphane Kastenbaum

International Partner: Mitsubishi Electric R&D Europe

Duration: 2015 - 2018

Abstract: The goal of this doctoral project is to verify and build cyber-physical systems (CPSs) with a correct-by-construction approach in order to validate system requirements against the two facets of the cyber and physical aspects of such designs. Our approach is based on components augmented with formal contracts that can be composed, abstracted or refined. It fosters the proof of system-level requirements by composing individual properties proved at component level. While semantically grounded, the tooling of this methodology should be usable by regular engineers (i.e. not proof theory specialists).

8.1.3. Mitsubishi Electric R&D Europe (2015-2019)

Title: Parallelism and modular proof in differential dynamic logic [1]

Inria principal investigator: Jean-Pierre Talpin, Simon Lunel

International Partner: Mitsubishi Electric R&D Europe

Duration: 2015 - 2018

[0] Towards verified programming of embedded devices. J.-P. Talpin, J.-J. Marty, S. Narayan, D. Stefan, R. Gupta. Design, Automation and Test in Europe (DATE'19). IEEE, 2019.

Abstract: The primary goal of this Ph.D. 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. International Initiatives

9.1.1. Inria International Labs

Sino-European Laboratory in Computer Science, Automation and Applied Mathematics
Associate Team involved in the International Lab:

9.1.1.1. CONVEX

Title: Compositional Verification of Cyber-Physical Systems

International Partner (Institution - Laboratory - Researcher):

CAS (China) - State Key Laboratory of Computer Science - Naijun Zhan

Start year: 2018

See also: http://convex.irisa.fr

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 verified integration of CPS components.

Inria@SiliconValley
Associate Team involved in the International Lab:

9.1.1.2. Composite

Title: Compositional System Integration

International Partners (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.

9.1.1.3. Inria International Chairs

IIC GUPTA Rajesh

Title: End-to-end system co-design

International Partner (Institution - Laboratory - Researcher):

University of California, San Diego (United States) - Rajesh Gupta

Duration: 2017 - 2021

Start year: 2017

9.1.1.4. Insa-Inria International Chair

Shuvra Bhattacharyya

Title: System design methodologies for real-time signal and information processing

International Partner (Institution - Laboratory - Researcher):

University of Maryland (United States) - Shuvra Bhattacharyya

Duration: 2018 - 2021

Start year: 2017

9.2. International Research Visitors

9.2.1. Visits of International Scientists

- Shuvra Bhattacharyya (UMD) visited project-team TEA and IETR in the context of his Insa-Inria Chair in May, July and December. He gave numerous talks and organized a workshop for the preparation of a European project proposal.
- Rajesh Gupta (UCSD) visited project-team TEA in the context of his Inria Chair in July and gave a seminar entitled: programming human spaces.
- Niki Vazou (IMDEA) visited project-team TEA in May and gave a presentation on her POPL'20 paper: "Liquidate your assets: reasoning about resource usage in Liquid Haskell".
- Yamine Ait Ameur (IRIT) visited project-team TEA in January on the occasion of Simon Lunel's Thesis defense.
- Naijun Zhan (ISCAS) visited project-team TEA in July, in the context of associate-project CONVEX.

- Delegates of the Sheng Yuan Honors College (BUAA) visited Inria-Irisa and Ecole Normale Supérieur de Rennes for the prospect of initiating an exchange program for graduate students, which will start in 2020.
- Zhang Bojun and Wang Zikai (BUAA) visited project-team TEA in July for an internship on verified modeling of blockchain protocols in Coq.
- Shenghao Yuan (NUAA) visited project-team TEA in July, in the context of associate-team CONVEX, and gave a presentation of the verified mini-Signal code generator developed at Nanhang University.

9.2.2. Visits to International Teams

Jean-Pierre Talpin visited UC San Diego in March, in the context of the associate-team Composite, and visited ISCAS, Beijing, in May and October, in the context of the associate-team CONVEX.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Selection

Jean-Pierre Talpin served in the program committee of the ACM LCTES'19, ACM SAC'19 and SCOPES'19 conferences.

Thierry Gautier reviewed articles for *Journal of Systems Architecture* (Elsevier).

10.1.2. Journal

Jean-Pierre Talpin is Associate Editor with the ACM Transactions for Embedded Computing Systems (TECS).

Thierry Gautier reviewed articles for *IEEE Access* and *Science of Computer Programming*.

10.1.3. Invited Talks

Jean-Pierre Talpin gave an invited presentation entitled "Towards verified programming of embedded devices" at DATE'19, Florence.

10.2. Teaching - Supervision - Juries

Jean-Pierre Talpin gave a one week graduate-level course at the Sheng Yuan Honors College, BUAA, entitled: "introduction to program verification".

Jean-Pierre Talpin co-supervises the PhD Theses of Stéphane Kastenbaum, Simon Lunel, Liangcong Zhang, Jean-Joseph Marty and Lucas Franceschino

Thierry Gautier served as external assessor for professor position application at Nankai University (China).

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] S. LUNEL.*Parallelism and modular proof in differential dynamic logic*, Université Rennes 1, January 2019, https://tel.archives-ouvertes.fr/tel-02102687

Articles in International Peer-Reviewed Journal

[2] 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", August 2019, vol. 13, n^o 4, p. 677-697 [*DOI :* 10.1007/s11704-017-6134-5], https://hal.inria.fr/hal-01411257

Invited Conferences

[3] J.-P. TALPIN, J.-J. MARTY, S. NARAYAN, D. STEFAN, R. GUPTA.*Towards verified programming of embedded devices*, in "DATE 2019 - 22nd IEEE/ACM Design, Automation and Test in Europe", Florence, Italy, IEEE, March 2019, p. 1445-1450 [*DOI :* 10.23919/DATE.2019.8715067], https://hal.inria.fr/hal-02193635

[4] H. ZHAN, Q. LIN, S. WANG, J.-P. TALPIN, X. XU, N. ZHAN.*Unified Graphical Co-Modelling of Cyber-Physical Systems using AADL and Simulink/Stateflow*, in "UTP 2019 - 7th International Symposium on Unifying Theories of Programming", Porto, Portugal, October 2019, p. 1-20 [*DOI :* 10.1007/978-3-030-31038-7_6], https://hal.inria.fr/hal-02193662

International Conferences with Proceedings

[5] S. LUNEL, S. MITSCH, B. BOYER, J.-P. TALPIN.*Parallel Composition and Modular Verification of Computer Controlled Systems in Differential Dynamic Logic*, in "FM 2019 - 23rd International Symposium on Formal Methods", Porto, Portugal, October 2019, p. 1-22, https://arxiv.org/abs/1907.02881 - Long version of an article accepted to the conference FM'19, https://hal.inria.fr/hal-02193642

[6] H. N. TRAN, A. HONORAT, J.-P. TALPIN, T. GAUTIER, L. BESNARD.*Efficient Contention-Aware Scheduling of SDF Graphs on Shared Multi-bank Memory*, in "ICECCS 2019 - 24th International Conference on Engineering of Complex Computer Systems", Hong Kong, China, IEEE, November 2019, p. 114-123 [*DOI :* 10.1109/ICECCS.2019.00020], https://hal.inria.fr/hal-02193639

Project-Team WIDE

The World Is DistributEd: Exploring the tension between scale and coordination

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

IN PARTNERSHIP WITH:
Université Rennes 1

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Distributed Systems and middleware

Table of contents

Project-Team WIDE

Creation of the Team: 2018 January 01, updated into Project-Team: 2018 June 01

Keywords:

Computer Science and Digital Science:

A1.2.5. - Internet of things
A1.2.9. - Social Networks
A1.3.2. - Mobile distributed systems
A1.3.3. - Blockchain
A1.3.4. - Peer to peer
A1.3.5. - Cloud
A1.3.6. - Fog, Edge
A2.1.7. - Distributed programming
A2.6.2. - Middleware
A2.6.3. - Virtual machines
A3.5.1. - Analysis of large graphs
A4.8. - Privacy-enhancing technologies
A7.1.1. - Distributed algorithms
A7.1.2. - Parallel algorithms
A7.1.3. - Graph algorithms
A9.2. - Machine learning

Other Research Topics and Application Domains:

B6.1.1. - Software engineering
B6.3.1. - Web
B6.3.5. - Search engines
B6.4. - Internet of things
B9.5.1. - Computer science
B9.5.6. - Data science

1. Team, Visitors, External Collaborators

Research Scientists

Davide Frey [Inria, Researcher, HDR]
George Giakkoupis [Inria, Researcher]
Anne-Marie Kermarrec [Inria, Senior Researcher, from Aug 2019, HDR]
Erwan Le Merrer [Inria, Advanced Research Position, HDR]

Faculty Members

Francois Taiani [Team leader, Univ de Rennes I, Professor, HDR]
David Bromberg [Univ de Rennes I, Professor, HDR]
Michel Raynal [Univ de Rennes I, Emeritus, HDR]

Technical Staff

Alex Auvolat Bernstein [Univ de Rennes I, Engineer, from Mar 2019 until Sep 2019]

PhD Students

Alex Auvolat Bernstein [Univ de Rennes I, PhD Student, from Oct 2019]
Loick Bonniot [Technicolor, PhD Student, granted by CIFRE]
Amaury Bouchra Pilet [Univ de Rennes I, PhD Student]
Florestan de Moor [Ecole normale supérieure de Rennes, PhD Student, from Sep 2019]
Quentin Dufour [Inria, PhD Student]
Louison Gitzinger [Univ de Rennes I, PhD Student]
Adrien Luxey [Univ de Rennes I, PhD Student]

Post-Doctoral Fellow

Noureddine Haouari [Univ de Rennes I, Post-Doctoral Fellow, from Mar 2019]

Visiting Scientists

Hasnaa Dyani [ENSIAS, Rabat, stagiaire, from Apr 2019 until Jun 2019]
Arsany Guirguis [EPFL, Lausanne, PhD student, from Jul 2019 until Sep 2019]
Lionel Kaboret [Joseph Ki-Zerbo University, Ouagadougou, PhD Student, from Sep 2019 until Oct 2019]
Mohamed Lechiakh [ENSIAS, Rabat, stagiaire, from Mar 2019 until May 2019]
Roberto Rodrigues Filho [Lancaster University, from Jul 2019 until Sep 2019]
Chaimaa Tarzi [ENSIAS, Rabat, stagiaire, from Apr 2019 until Jun 2019]

2. Overall Objectives

2.1. Overview

The long term goal of the WIDE team is to provide the practical tools and theoretical foundations required to address the scale, dynamicity, and uncertainty that constitute the foundations of modern distributed computer systems. In particular, we would like to **explore the inherent tension between scalability and coordination guarantees**, and develop novel techniques and paradigms that are adapted to the rapid and profound changes impacting today's distributed systems, both in terms of the application domains they support and the operational constraints they must meet.

These changes are particularly visible in three key areas related to our research: *(i)* planetary-scale information systems, *(ii)* personalized services, and *(iii)* new forms of social applications (e.g. in the field of the sharing economy).

2.2. Planetary-Scale Geo-Distributed Systems

Modern large-scale systems often encompass thousands of server nodes, hosted in tens of datacenters distributed over several continents. To address the challenges posed by such systems, alternative distributed architectures are today emerging that emphasize *decentralized* and *loosely coupled* interactions. This evolution can be observed at multiple levels of an application's distributed stack: the growing interest, both practical and theoretical, for weak consistency models is such an example. In spite of their potential counter-intuitive behaviors, weakly consistent data-structures allow developers to trade strict coordination guarantees for the ability to deliver a reactive and scalable service even when hit by arbitrary network delays or system partitions. At a higher, more architectural level, similar motivations explain the push for *micro-services* on the server side of on-line applications and the growth of rich *browser-based programming technologies* on their client side. Micro services help development teams decompose complex applications into a set of simpler and loosely-connected distributed services. In a parallel evolution, modern browsers embark increasingly powerful networking APIs such as WebRTC. These APIs are prompting a fresh rethink of the typical distribution of capabilities between servers and clients. This is likely to lead to more services and computations being offloaded to browsers, in particular within hybrid architectures. The above evolutions, away from tightly synchronized and monolithic deployments towards heterogeneous, composite and loosely coordinated distributed systems, raise a number of difficult challenges at the crossroad of theoretical distributed algorithms, system architecture, and programming frameworks. One of these challenges pertains to the growing complexity arising from these systems: as richer and more diverse services are being composed to construct whole applications, individual developers can only hope to grasp parts of the resulting systems. Similarly, weak consistency models and loose coordination mechanisms tend to lead to counter-intuitive behaviors, while only providing weak overall guarantees. This lack of systematic guarantees and understandability make it harder for practitioners to design, deploy, and validate the distributed systems they produce, leading to rising costs and high entry barriers.

In order to address these challenges, we argue that modern-day distributed systems require new principled algorithms, approaches, and architectural patterns able to provide sound foundations to their development while guaranteeing robust service guarantees, thus lowering the cost of their development and maintenance, increasing their reliability, and rendering them technically approachable to a wider audience.

2.3. Highly Personalized On-Line Services

Ever increasing volumes of data are being produced and made available from a growing number of sources (Internet of Things sensors, open data repositories, user-generated content services). As a result, digital users find it increasingly difficult to face the data deluge they are subjected to without additional help. This difficulty has fueled the rise of notification solutions over traditional search, in order to push few but relevant information items to users rather than leave them to sieve through a large mass of non-curated data. To provide such personalized services, most companies rely today on centralized or tightly coupled systems hosted in data centers or in the cloud. These systems use advanced data-mining and machine learning techniques to deliver enhanced, personalized, services to users and companies, and often exploit highly parallelized data analytics frameworks such as Spark, and Flink.

Selecting the best information for a user in order to provide a personalized experience requires however to gather enough information about this user, which raises a number of important technical challenges and privacy protection issues. More precisely, this concentration poses strong risks to the privacy of users, and limits the scope of personalization to tightly integrated datasets. The use of large monolithic infrastructures also limits the use of machine learning and personalization to situations in which data is fully available to the organization managing the underlying computing infrastructure. This set-up prevents for instance cases in which sensitive data may not be shared freely, but might be of mutual interest to several independent participants in order to construct common machine learning models usable by all. Such situations occur for instance in the context of the mining of health-records by independent health-organizations, or in the collective harnessing of individual on-line profiles for personalization purpose by private users.

Alternative decentralized approaches that eschew the need for a central all-encompassing authority holds the promise of delivering knowledge while protecting individual participants. Constructing such systems requires however to address the inherent tension between the need to limit sensitive individual leaks, while maximizing collectively gained insights. Answering this tension calls on techniques and approaches from distributed systems, information theory, security, and randomized processes, making it a rich and dense research area, with a high impact potential. The problem of distributed privacy in a digital interconnected age further touches on interdisciplinary questions of Law, Sociology and Public Policy, which we think can only be explored in collaboration with colleagues from these fields.

2.4. Social Collaboration Platforms

On-line social networks have had a fundamental and lasting impact on the Internet. In recent years, numerous applications have appeared that go beyond the services originally provided by “pure” on-line social networks, such as posting messages or maintaining on-line “friendship” links. These new applications seek to organize and coordinate users, often in the context of the sharing economy, for instance in order to facilitate car-sharing (e.g. BlaBla car, https://www.blablacar.com/), short-term renting (e.g. AirBnB, https://www.airbnb.com/), and peer-to-peer financial services (e.g. Lending Club, https://www.lendingclub.com/). Some systems, such as Bitcoin or Ethereum, have given rise to new distributed protocols combining elements of cryptography and distribution that are now largely discussed in the research community, and have attracted the attention of policy makers and leading financial actors.

The challenges faced by such social applications blend in many ways issues already discussed in the two previous subsections and cast them in an application-driven context. These social collaboration platforms require mechanisms that go beyond pure message propagation, with stricter consistency and robustness guarantees. Because they involve connected users, these applications must provide usable solutions, in particular in terms of latency and availability. At the same time, because they manipulate real-world transactions and objects (money, cars, accommodations) they must also provide a high level of consistency and guarantees. Many of these applications further operate at a planetary scale, and therefore also face stark scalability issues, that make them highly interesting case studies to investigate innovative architectures combining decentralized and centralized elements.

Formalizing and characterizing the needs and behaviors of these new applications seems particularly interesting in order to provide the fertile ground for new systems and novel theoretical work. The area of social applications also offers avenues for knowledge transfer and societal impact, along two dimensions. First, practical and usable approaches, back by a deep understanding of the foundation of distribution and coordination, are likely to find applications in future systems. Second, developers of complex social applications are often faced with a lack of robust scalable services [0] that can be easily exploited to harness the latest understanding of large-scale distributed coordination. We therefore think these applications offer an opportunity to design and deliver modular reusable bricks that can be easily appropriated by a large population of innovative developers without requiring the level of deep understanding usually necessary to implement these solutions from scratch. Providing such reusable bricks is however difficult, as many interesting formal properties are not composable, and a unified composable theory of distributed systems still need to be fully articulated.

3. Research Program

3.1. Overview

In order to progress in the four fields described above, the WIDE team is developing a research program which aims to **help developers control and master the inherent uncertainties and performance challenges brought by scale and distribution**.

More specifically, our program revolves around four key challenges.

- Challenge 1: Designing Hybrid Scalable Architectures,
- Challenge 2: Constructing Personalizable Privacy-Aware Distributed Systems,
- Challenge 3: Understanding Controllable Network Diffusion Processes,
- Challenge 4: Systemizing Modular Distributed Computability and Efficiency.

These four challenges have in common **the inherent tension between coordination and scalability in large-scale distributed systems**: strong coordination mechanisms can deliver strong guarantees (in terms of consistency, agreement, fault-tolerance, and privacy protection), but are generally extremely costly and inherently non-scalable if applied indiscriminately. By contrast, highly scalable coordination approaches (such as epidemic protocols, eventual consistency, or self-organizing overlays) perform much better when the size of a system increases, but do not, in most cases, provide any strong guarantees in terms of consistency or agreement.

The above four challenges explore these tensions from *four complementary angles*: from an architectural perspective (Challenge 1), from the point of view of a fundamental system-wide guarantee (privacy protection, Challenge 2), looking at one universal scalable mechanism (network diffusion, Challenge 3), and considering the interplay between modularity and computability in large-scale systems (Challenge 4). These four challenges range from practical concerns (Challenges 1 and 2) to more theoretical questions (Challenrges 3 and 4),

[0] The repeated debugging of MongoDB's replication algorithm (e.g. see https://aphyr.com/posts/338-jepsen-mongodb-3-4-0-rc3) is a telling illustration of the difficulties encountered by development teams when building such platforms.

yet present *strong synergies* and *fertile interaction points*. E.g. better understanding network diffusion (Challenge 3) is a key enabler to develop more private decentralized systems (Challenge 2), while the development of a theoretically sound modular computability hierarchy (Challenge 4) has a direct impact on our work on hybrid architectures (Challenge 1).

3.2. Hybrid Scalable Architectures

The rise of planetary-scale distributed systems calls for novel software and system architectures that can support user-facing applications while scaling to large numbers of devices, and leveraging established and emerging technologies. The members of WIDE are particularly well positioned to explore this avenue of research thanks to their experience on de-concentrated architectures combining principles from both decentralized peer-to-peer [48], [58] systems and hybrid infrastructures (i.e. architectures that combines centralized or hierarchical elements, often hosted in well-provisioned data-centers, and a decentralized part, often hosted in a peer-to-peer overlay) [52]. In the short term, we aim to explore two axes in this direction: browser-based communication, and micro services.

3.2.1. Browser-based fog computing

The dramatic increase in the amount of data being produced and processed by connected devices has led to paradigms that seek to decentralize the traditional cloud model. In 2011 Cisco [49] introduced the vision of *fog computing* that combines the cloud with resources located at the edge of the network and in between. More generally, the term *edge computing* has been associated with the idea of adding edge-of-the network storage and computation to traditional cloud infrastructures [44].

A number of efforts in this directions focus on specific hardware, e.g. fog nodes that are responsible for connected IoT devices [50]. However, many of today's applications run within web browsers or mobile phones. In this context, the recent introduction of the WebRTC API, makes it possible for browsers and smartphones to exchange directly between each other, enabling mobile, or browser-based decentralized applications. Maygh [72], for example, uses the WebRTC API to build a decentralized Content Delivery Network that runs solely on web browsers. The fact that the application is hosted completely on a web server and downloaded with enabled websites means that webmasters can adopt the Content Delivery Network (CDN) without requiring users to install any specific software.

For us, the ability of browsers to communicate with each other using the WebRTC paradigm provides a novel playground for new programming models, and for a *browser-based fog architecture* combining both a centralized, cloud-based part, and a decentralized, browser-supported part.

This model offers tremendous potential by making edge-of-the-network resources available through the interconnection of web-browsers, and offers new opportunities for the protection of the personal data of end users. But consistently engineering browser-based components requires novel tools and methodologies.

In particular, WebRTC was primarily designed for exchanging media and data between two browsers in the presence of a coordinating server. Its complex mechanisms for connection establishment make many of the existing peer-to-peer protocols inefficient. To address this challenge, we plan to consider two angles of attack. First, we plan to design novel protocols that take into account the specific requirements set by this new technology. Second, we envisage to investigate variants of the current WebRTC model with cheaper connection-establishment protocols, in order to provide lower delays and bandwidth consumption in large-scale browser-based applications.

We also plan to address the trade-offs associated with hybrid browser-cloud models. For example, when should computation be delegated to browsers and when should it be executed on the cloud in order to maximize the quality of service? Or, how can a decentralized analytics algorithms operating on browser-based data complement or exploit the knowledge built by cloud-based data analytics solutions?

3.2.2. Emergent micro-service deployment and management

Micro-services tend to produce fine-grained applications in which many small services interact in a loosely coupled manner to produce a wide range of services within an organization. Individual services need to evolve independently of each other over time without compromising the availability of the overall application. Lightweight isolation solutions such as containers (Docker, ...), and their associated tooling ecosystem (e.g. Google's Borg [71], Kubernetes [47]) have emerged to facilitate the deployment of large-scale micro-service-based applications, but only provide preliminary solutions for key concerns in these systems, which we would like to investigate and extend.

Most of today's on-line computer systems are now too large to evolve in monolithic, entirely pre-planned ways. This applies to very large data centres, for example, where the placement of virtual machines to reduce heating and power consumption can no longer be treated using top-down exhaustive optimisation approaches beyond a critical size. This is also true of social networking applications, where different mechanisms—e.g. to spread news notifications, or to recommend new contacts—must be adapted to the different sub-communities present in the system.

To cope with the inherent complexity of building complex loosely-coupled distributed systems while fostering and increasing efficiency, maintainability, and scalability, we plan to study how novel programming techniques based on declarative programming, components and epidemic protocols can help design, deploy, and maintain self-adaptive structures (e.g. placement of VM) and mechanisms (e.g. contact recommendations) that are optimized to the local context of very large distributed systems. To fulfill this vision, we plan to explore a three-pronged strategy to raise the level of programming abstraction offered to developers.

- First, we plan to explore the use of high-level domain-specific languages (DSL) to declare how large-scale topologies should be achieved, deployed, and maintained. Our vision is a declarative approach to describe how to combine, deploy and orchestrate micro-services in an abstract manner thus abstracting away developers from the underlying cloud infrastructures, and from the intricacies involved in writing low-level code to build a large-scale distributed application that scales. With this effort, we plan notably to directly support the twin properties of *emergence* (the adaptation "from within") and *differentiation* (the possibility from parts of the system to diverge while still forming a whole). Our central objective is to search for principled programming constructs to support these two capabilities using a modular and incremental software development approach.
- On a second strand of work, we plan to investigate how unikernels enable smaller footprints, more optimization options, and faster boot times for micro-services. Isolating micro-services into VMs is not the most adequate approach as it requires the use of hypervisors, or virtual machine monitors (VMMs), to virtualize hardware resources. VMMs are well known to be heavyweight with both boot and run time overheads that may have a strong impact on performances. Unikernels seem to offer the right balance between performance and flexibility to address this challenge. One of the key underlying challenges is to compile directly the aforementioned provided DSL to a dedicated and customized machine image, ready to be deployed directly on top of a large set of bare metal servers.
- Depending on the workload it is subjected to, and the state of its execution environment (network, VMs), a large-scale distributed application may present erratic or degraded performance that is hard to anticipate and plan for. There is therefore a strong need to adapt dynamically the way resources are allocated to a running application. We would like to study how the DSL approach we envisage can be extended to enable developers to express orchestration algorithms based on machine learning algorithms.

3.3. Personalizable Privacy-Aware Distributed Systems

On-line services are increasingly moving towards an in-depth analysis of user data, with the objective of providing ever better personalization. But in doing so, personalized on-line services inevitably pose risks to the privacy of users. Eliminating, or even reducing these risks raises important challenges caused by the inherent trade-off between the level of personalization users wish to achieve, and the amount of information they are willing to reveal about themselves (explicitly or through the many implicit sources of digital information such as smart homes, smart cars, and IoT environments).

At a general level, we would like to address these challenges through protocols that can provide access to unprecedented amounts of data coming from sensors, users, and documents published by users, while protecting the privacy of individuals and data sources. To this end, we plan to rely on our experience in the context of distributed systems, recommender systems, and privacy, as well as in our collaborations with experts in neighboring fields such as machine learning, and security. In particular, we aim to explore different privacy-utility tradeoffs that make it possible to provide differentiated levels of privacy guarantees depending on the context associated with data, on the users that provide the data, and on those that access it. Our research targets the general goal of privacy-preserving decentralized learning, with applications in different contexts such as user-oriented applications, and the Internet-of-Things (IoT).

3.3.1. Privacy-preserving decentralized learning

Personalization and recommendation can be seen as a specific case of general machine learning. Production-grade recommenders and personalizers typically centralize and process the available data in one location (a data-center, a cloud service). This is highly problematic, as it endangers the privacy of users, while hampering the analysis of datasets subject to privacy constraints that are held by multiple independent organizations (such as health records). A decentralized approach to machine learning appears as a promising candidate to overcome these weaknesses: if each user or participating organization keeps its data, while only exchanging gradient or model information, privacy leaks seem less likely to occur.

In some cases, decentralized learning may be achieved through relatively simple adaptations of existing centralized models, for instance by defining alternative learning models that may be more easily decentralized. But in all cases, processing growing amounts of information calls for high-performance algorithms and middleware that can handle diverse storage and computation resources, in the presence of dynamic and privacy-sensitive data. To reach this objective, we will therefore leverage our work in distributed and privacy-preserving algorithms and middleware [51], [53], [54] as well as the results of our work on large-scale hybrid architectures in Objective 1.

3.3.2. Personalization in user-oriented applications

As a first application perspective, we plan to design tools that exploit decentralized analytics to enhance user-centric personalized applications. As we observed above, such applications exhibit an inherent trade-off between personalization quality and privacy preservation. The most obvious goal in this direction consists in designing algorithms that can achieve high levels of personalization while protecting sensitive user information. But an equally important one consists in personalizing the trade-off itself by adapting the quality of the personalization provided to a user to his/her willingness to expose information. This, like other desirable behaviors, appears at odds with the way current systems work. For example, a user of a recommender system that does not reveal his/her profile information penalizes other users causing them to receive less accurate recommendations. We would like to mitigate this situation by means of protocols that reward users for sharing information. On the one hand, we plan to take inspiration from protocols for free-riding avoidance in peer-to-peer systems [55], [60]. On the other hand, we will consider blockchains as a tool for tracking and rewarding data contributions. Ultimately, we aim at enabling users to configure the level of privacy and personalization they wish to experience.

3.3.3. Privacy preserving decentralized aggregation

As a second setting we would like to consider target applications running on constrained devices like in the Internet-of-Things (IoT). This setting makes it particularly important to operate on decentralized data in a light-weight privacy-preserving manner, and further highlights the synergy between this objective and Objective 1. For example, we plan to provide data subjects with the possibility to store and manage their data locally on their own devices, without having to rely on third-party managers or aggregators, but possibly storing less private information or results in the cloud. Using this strategy, we intend to design protocols that enable users themselves, or third-party companies to query distributed data in aggregate form, or to run data analytics processes on a distributed set of data repositories, thereby gathering knowledge without violating the privacy of other users. For example, we have started working on the problem of computing an aggregate function over a subset of the data in a distributed setting. This involves two major steps: selection and aggregation. With respect to selection, we envision defining a decentralized data-selection operation that can apply a selection predicate without violating privacy constraints. With respect to aggregation, we will continue our investigation of lightweight protocols that can provide privacy with limited computational complexity [45].

3.4. Network Diffusion Processes

Social, biological, and technological networks can serve as conduits for the spread of ideas, trends, diseases, or viruses. In social networks, rumors, trends and behaviors, or the adoption of new products, spread from person to person. In biological networks, diseases spread through contact between individuals, and mutations spread from an individual to its offsprings. In technological networks, such as the Internet and the power grid, viruses and worms spread from computer to computer, and power failures often lead to cascading failures. The common theme in all the examples above is that the rumor, disease, or failure starts out with a single or a few individual nodes, and propagates through the network, from node to node, to reach a potentially much larger number of nodes.

These types of *network diffusion processes* have long been a topic of study in various disciplines, including sociology, biology, physics, mathematics, and more recently, computer science. A main goal has been to devise mathematical models for these processes, describing how the state of an individual node can change as a function of the state of its neighbors in the network, and then analyse the role of the network structure in the outcome of the process. Based on our previous work, we would like to study to what extent one can affect the outcome of the diffusion process by controlling a small, possibly carefully selected fraction of the network.

For example, we plan to explore how we may 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).

Our goal is to provide a framework for a systematic and rigorous study of these problems. We will consider several standard diffusion models and extensions of them, including models from mathematical sociology, mathematical epidemiology, and interacting particle systems. We will consider existing and new variants of spread maximization/limitation problems, and will provide (approximation) algorithms or show negative (inapproximability) results. In case of negative results, we will investigate general conditions that make the problem tractable. We will consider both general network topologies and specific network models, and will relate the efficiency of solutions to structural properties of the topology. Finally, we will use these insights to engineer new network diffusion processes for efficient data dissemination.

3.4.1. Spread maximization

Our goal is in particular to study spread maximization in a broader class of diffusion processes than the basic independent cascade (IC) and linear threshold (LT) models of influence [64], [65], [66] that have been studied in this context so far. This includes the *randomized rumor spreading (RS)* model for information dissemination [57], *biased* versions of the *voter model* [61] modelling influence, and the (graph-based) *Moran processes* [68] modelling the spread of mutations. We would like to consider several natural versions of the spread maximization problem, and the relationships between them. For these problems we will use the greedy algorithm and the submodularity-based analytical framework of [64], and will also explore new approaches.

3.4.2. Immunization optimization

Conversely we would also like to explore immunization optimization problems. Existing works on these types of problem assume a *perfect-contagion* model, i.e., once a node gets infected, it deterministically infects all its non-immunized neighbors. We plan to consider various diffusion processes, including the standard *susceptible–infected* (SI), *susceptible–infected–recovered* (SIR) and *susceptible–infected–susceptible* (SIS) epidemic models, and explore the extent to which results and techniques for the perfect-contagion model carry over to these probabilistic models. We will also investigate whether techniques for spread maximization could be applied to immunization problems.

Some immunization problems are known to be hard to approximate in general graphs, even for the perfect-contagion model, e.g., the fixed-budget version of the fire-fighter problem cannot be approximated to any $n^{1-\epsilon}$ factor [46]. This strand of work will consider restricted graph families, such as trees or graphs of small treewidth, for such problems. In addition, for some immunization problems, there is a large gap between the best known approximation algorithm and the best known inaproximability result, and we would like to make progress in reducing these gaps.

3.5. Systemizing Modular Distributed Computability and Efficiency

The applications and services envisaged in Objectives 1 and 2 will lead to increasingly complex and multifaceted systems. Constructing these novel hybrid and decentralized systems will naturally push our need to understand distributed computing beyond the current state of the art. These trends therefore demand research efforts in establishing sound theoretical foundations to allow everyday developers to master the design, properties and implementation of these systems. We plan to investigate these foundations along two directions: first by studying novel approaches to some fundamental problems of *mutual exclusion and distributed coordination*, and second by exploring how we can build a *comprehensive and modular framework* capturing the foundations of *distributed computation*.

3.5.1. Randomized algorithm for mutual exclusion and coordination

To exploit the power of massive distributed applications and systems (such as those envisaged in Objectives 1 and 2) or multiple processors, algorithms must cope with the scale and asynchrony of these systems, and their inherent instability, e.g., due to node, link, or processor failures. Our goal is to explore the power and limits of randomized algorithms for large-scale networks of distributed systems, and for shared memory multi-processor systems, in effect providing fundamental building blocks to the work envisioned in Objectives 1 and 2.

For shared memory systems, randomized algorithms have notably proved extremely useful to deal with asynchrony and failures. Sometimes probabilistic algorithms provide the only solution to a problem; sometimes they are more efficient; sometimes they are simply easier to implement. We plan to devise efficient algorithms for some of the fundamental problems of shared memory computing, such as mutual exclusion, renaming, and consensus.

In particular, looking at the problem of *mutual exclusion*, it is desirable that mutual exclusion algorithms be *abortable*. This means that a process that is trying to lock the resource can abort its attempt in case it has to wait too long. Abortability is difficult to achieve for mutual exclusion algorithms. We will try to extend our algorithms for the *cache-coherent* (CC) and the *distributed shared memory* (DSM) model in order to make them abortable, while maintaining expected constant *Remote Memory References* (RMRs) complexity, under optimistic system assumptions. In order to achieve this, the algorithm will use strong synchronization primitives, called compare-and-swap objects. As part of our collaboration with the University of Calgary, we will work on implementing those objects from registers in such a way that they also allow aborts. Our goal is to build on existing non-abortable implementations [59]. We plan then later to use these objects as building blocks in our mutual exclusion algorithm, in order to make them work even if the system does not readily provide such primitives.

We have also started working on blockchains, as these represent a new and interesting trade-off between probabilistic guarantees, scalability, and system dynamics, while revisiting some of the fundamental questions and limitations of consensus in fault-prone asynchronous systems.

3.5.2. Modular theory of distributed computing

Practitioners and engineers have proposed a number of reusable frameworks and services to implement specific distributed services (from Remote Procedure Calls with Java RMI or SOAP-RPC, to JGroups for group communication, and Apache Zookeeper for state machine replication). In spite of the high conceptual and practical interest of such frameworks, many of these efforts lack a sound grounding in distributed computation theory (with the notable exceptions of JGroups and Zookeeper), and often provide punctual and partial solutions for a narrow range of services. We argue that this is because we still lack a generic framework that unifies the large body of fundamental knowledge on distributed computation that has been acquired over the last 40 years.

To overcome this gap we would like to develop 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. This research vision 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.

To progress on this vision, we plan in the near future to investigate, from a distributed software point of view, the impact due to failures and asynchrony on the layered architecture of distributed computing systems. A first step in this direction will address the notions of *message adversaries* (introduced a long time ago in [70]) and *process adversaries* (investigated in several papers, e.g. [69], [56], [62], [63], [67]). The aim of these notions is to consider failures, not as "bad events", but as part of the normal behavior of a system. As an example, when considering round-based algorithms, a message adversary is a daemon which, at every round, is allowed to suppress some messages. The aim is then, given a problem P, to find the strongest adversary under which P can be solved ("strongest" means here that giving more power to the adversary makes the problem impossible to solve). This work will allow us to progress in terms of general *layered* theory of distributed computing, and allow us to better *map* distributed computing models and their relations, in the steps of noticeable early efforts in this direction [69], [43].

4. Highlights of the Year

4.1. Highlights of the Year

4.1.1. Awards

Florestan De Moor is the recipient of the "Prix National Jeunes André Blanc-Lapierre 2019" from the SEE society (Société de l'électricité, de l'électronique et des technologies de l'information et de la communication), for his work during his master thesis .

During the SRDS 2019 conference which took place in Lyon, France, from October 1st to 4th, Michel Raynal received an Outstanding Career Award for his contributions to distributed systems and algorithms.

BEST PAPERS AWARDS :

[42]

F. DE MOOR.*A Biclustering Approach to Recommender Systems*, University of Rennes 1, June 2019, p. 1-46, https://hal.inria.fr/hal-02369708

5. New Software and Platforms

5.1. WebGC

Web-based Gossip Communication

KEYWORDS: Epidemic protocols - Gossip protocols - Peer-to-peer - Web - Personalized systems - Decentralized architectures - Recommendation systems - WebRTC - 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. YALPS

Keywords: Traffic-shaping - Nat traversal - Experimentation - Peer-to-peer - Simulator - 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/

5.3. KIFF

KIFF: An impressively fast and efficient JAVA library for KNN construction

Keyword: KNN

Functional Description: This package implements the KIFF algorithm reported in [1]. KIFF is a generic, fast and scalable K-Nearest-Neighbor graph construction algorithm. This algorithm connects each object to its k most similar counterparts, according to a given similarity metric. In term of comparison, this package implements also HYREC [2] and NN-DESCENT [3]. The standalone program implements cosine similarity only, however this library supports arbitrary similarity measures.

[1] Antoine Boutet, Anne-Marie Kermarrec, Nupur Mittal, Francois Taiani. Being prepared in a sparse world: the case of KNN graph construction. ICDE 2016, Finland.

- Partner: LIRIS
- Contact: Antoine Boutet

6. New Results

6.1. Recommender Systems

6.1.1. A Biclustering Approach to Recommender Systems

Participants: Florestan de Moor, Davide Frey.

Recommendation systems are a core component of many e-commerce industries and online services since they ease the discovery of relevant products. Because catalogs are huge, it is impossible for an individual to manually search for an item of interest, hence the need for some automatic filtering process. Many approaches exist, from content-based ones to collaborative filtering that include neighborhood and model-based techniques. Despite these intensive research activities, numerous challenges remain to be addressed, particularly under real-time settings or regarding privacy concerns, which motivates further work in this area. We focus on techniques that rely on biclustering, which consists in simultaneously building clusters over the two dimensions of a data matrix. Although it was little considered by the recommendation system community, it is a well-known technique in other domains such as genomics. In work [42] we present the different biclustering-based approaches that were explored. We then are the first to perform an extensive experimental evaluation to compare these approaches with one another, but also with the current state-of-the-art techniques from the recommender field. Existing evaluations are often restrained to a few algorithms and consider only a limited set of metrics. We then expose a few ideas to improve existing approaches and address the current challenges in the design of highly efficient recommendation algorithms, along with some preliminary results.

This work was done in collaboration with Antonio Mucherino (University of Rennes 1).

6.1.2. Unified and Scalable Incremental Recommenders with Consumed Item Packs

Participant: Erwan Le Merrer.

Recommenders personalize the web content by typically using collaborative filtering to relate users (or items) based on explicit feedback, e.g., ratings. The difficulty of collecting this feedback has recently motivated to consider implicit feedback (e.g., item consumption along with the corresponding time). In this work [39], we introduce the notion of consumed itempack (CIP) which enables to link users (or items) based on their implicit analogous consumption behavior. Our proposal is generic, and we show that it captures three novel implicit recommenders: a user-based (CIP-U), an item-based (CIP-I),and a word embedding-based (DEEPCIP), as well as a state-of-art technique using implicit feedback (FISM). We show that our recommenders handle incremental updates incorporating freshly consumed items. We demonstrate that all three recommenders provide a recommendation quality that is competitive with state-of-the-art ones, including one incorporating both explicit and implicit feedback

This work was done in collaboration with Rachid Guerraoui (EPFL), Rhicheek Patra (Oracle) and Jean-Ronan Vigouroux (Technicolor).

6.2. Systems for the Support of Privacy

6.2.1. Robust Privacy-Preserving Gossip Averaging

Participants: Amaury Bouchra-Pilet, Davide Frey, François Taïani.

This contribution aims to address the privacy risks inherent in decentralized systems by considering the emblematic problem of privacy-preserving decentralized averaging. In particular, we propose a novel gossip protocol that exchanges noise for several rounds before starting to exchange actual data. This makes it hard for an honest but curious attacker to know whether a user is transmitting noise or actual data. Our protocol and analysis do not assume a lock-step execution, and demonstrate improved resilience to colluding attackers. In a paper, publishing this work at SSS 2019 [26], we prove the correctness of this protocol as well as several privacy results. Finally, we provide simulation results about the efficiency of our averaging protocol.

6.2.2. A Collaborative Strategy for Mitigating Tracking through Browser Fingerprinting.

Participants: David Bromberg, Davide Frey, Alejandro Gomez-Boix.

Browser fingerprinting is a technique that collects information about the browser configuration and the environment in which it is running. This information is so diverse that it can partially or totally identify users online. Over time, several countermeasures have emerged to mitigate tracking through browser fingerprinting. However, these measures do not offer full coverage in terms of privacy protection, as some of them may introduce inconsistencies or unusual behaviors, making these users stand out from the rest.

In this work, we address these limitations by proposing a novel approach that minimizes both the identifiability of users and the required changes to browser configuration. To this end, we exploit clustering algorithms to identify the devices that are prone to share the same or similar fingerprints and to provide them with a new non-unique fingerprint. We then use this fingerprint to automatically assemble and run web browsers through virtualization within a docker container. Thus all the devices in the same cluster will end up running a web browser with an indistinguishable and consistent fingerprint.

We carried out this work in collaboration with Benoit Baudry from KTH Sweden and published our results at the 2019 Moving-Target Defense Workshop [30].

6.3. Distributed Algorithms

6.3.1. One for All and All for One:Scalable Consensus in a Hybrid Communication Model

Participant: Michel Raynal.

This work [34] addresses consensus in an asynchronous model where the processes are partitioned into clusters. Inside each cluster, processes can communicate through a shared memory, which favors efficiency. Moreover, any pair of processes can also communicate through a message-passing communication system, which favors scalability. In such a "hybrid communication" context, the work presents two simple binary consensus algorithms (one based on local coins,the other one based on a common coin). These algorithms are straightforward extensions of existing message-passing randomized round-based consensus algorithms. At each round, the processes of each cluster first agree on the same value (using an underlying shared memory consensus algorithm), and then use a message-passing algorithm to converge on the same decided value. The algorithms are such that, if all except one processes of a cluster crash, the surviving process acts as if all the processes of its cluster were alive (hence the motto "one for all and all for one"). As a consequence, the hybrid communication model allows us to obtain simple, efficient, and scalable fault-tolerant consensus algorithms. As an important side effect, according to the size of each cluster, consensus can be obtained even if a majority of processes crash.

This work was done in collaboration with Jiannong Cao (Polytechnic University, Hong Kong).

6.3.2. Optimal Memory-Anonymous Symmetric Deadlock-Free Mutual Exclusion

Participant: Michel Raynal.

The notion of an anonymous shared memory (recently introduced in PODC 2017) considers that processes use different names for the same memory location. Hence, there is permanent disagreement on the location names among processes. In this context, the PODC paper presented -among other results- a symmetric deadlock-free mutual exclusion (mutex) algorithm for two processes and a necessary condition on the size m of the anonymous memory for the existence of a symmetric deadlock-free mutex algorithm in an n-process system. This work [22] answers several open problems related to symmetric deadlock-free mutual exclusion in an n-process system where the processes communicate through m registers. It first presents two algorithms. The first considers that the registers are anonymous read/write atomic registers and works for any m greater than 1 and belonging to the set M(n). It thus shows that this condition on m is both necessary and sufficient. The second algorithm considers anonymous read/modify/write atomic registers. It assumes that $m \in M(n)$. These algorithms differ in their design principles and their costs (measured as the number of registers which must contain the identity of a process to allow it to enter the critical section). The work also shows that the condition

$m \in M(n)$ is necessary for deadlock-free mutex on top of anonymous read/modify/write atomic registers. It follows that, when m > 1, $m \in M(n)$ is a tight characterization of the size of the anonymous shared memory needed to solve deadlock-free mutex, be the anonymous registers read/write or read/modify/write.

This work was done in collaboration with Zahra Aghazadeh (University of Calgary), Damien Imbs (LIS, Université d'Aix-Marseille,CNRS, Université de Toulon), Gadi Taubenfeld (The Interdisciplinary Center of Herzliya) and Philipp Woelfel (University of Calgary).

6.3.3. Merkle Search Trees

Participants: Alex Auvolat, François Taïani.

Most recent CRDT (Conflict-free Replicated Data Type) techniques rely on a causal broadcast primitive to provide guarantees on the delivery of operation deltas. Such a primitive is unfortunately hard to implement efficiently in large open networks, whose membership is often difficult to track. As an alternative, we argue that pure state-based CRDTs can be efficiently implemented by encoding states as specialized Merkle trees, and that this approach is well suited to open networks where many nodes may join and leave. Indeed, Merkle trees enable efficient remote comparison and reconciliation of data sets, which can be used to implement the CRDT merge operator between two nodes without any prior information. This approach also does not require vector clock information, which would grow linearly with the number of participants.

At the core of our contribution [24] lies a new kind of Merkle tree, called Merkle Search Tree (MST), that implements a balanced search tree while maintaining key ordering. This latter property makes it particularly efficient in the case of updates on sets of sequential keys, a common occurrence in many applications. We use this new data structure to implement a distributed event store, and show its efficiency in very large systems with low rates of updates. In particular, we show that in some scenarios our approach is able to achieve both a 66% reduction of bandwidth cost over a vector-clock approach, as well as a 34% improvement in consistency level. We finally suggest other uses of our construction for distributed databases in open networks.

6.3.4. Dietcoin: Hardening Bitcoin Transaction Verification Process For Mobile Devices

Participants: Davide Frey, François Taïani.

Distributed ledgers are among the most replicated data repositories in the world. They offer data consistency, immutability, and auditability, based on the assumption that each participating node locally verifies their entire content. Although their content, currently extending up to a few hundred gigabytes, can be accommodated by dedicated commodity hard disks, downloading it, processing it, and storing it in general-purpose desktop and laptop computers can prove largely impractical. Even worse, this becomes a prohibitive restriction for smartphones, mobile devices, and resource-constrained IoT devices.

We thus proposed Dietcoin, a Bitcoin protocol extension that allows nodes to perform secure local verification of Bitcoin transactions with small bandwidth and storage requirements. We carried out an extensive evaluation of the features of Dietcoin that are important for today's cryptocurrency and smart-contract systems, but are missing in the current state-of-the-art. These include (i) allowing resource-constrained devices to verify the correctness of selected blocks locally without having to download the complete ledger; (ii) enabling devices to join a blockchain quickly yet securely, dropping bootstrap time from days down to a matter of seconds; (iii) providing a generic solution that can be applied to other distributed ledgers secured with Proof-of-Work. We showcased our results in a demo at VLDB 2019 [15], and we are currently preparing a full paper submission.

We carried out this work in collaboration with Pierre-Louis Roman, now at University of Lugano (Switzerland), as well as with Mark Makke from Vrije Universiteit, Amsterdam (the Netherlands), and Spyros Voulgaris from Athens University of Economics and Business (Greece).

6.3.5. Byzantine-Tolerant Set-Constrained Delivery Broadcast

Participants: Alex Auvolat, François Taïani, Michel Raynal.

Set-Constrained Delivery Broadcast (SCD-broadcast), recently introduced at ICDCN 2018, is a high-level communication abstraction that captures ordering properties not between individual messages but between sets of messages. More precisely, it allows processes to broadcast messages and deliver sets of messages, under the constraint that if a process delivers a set containing a message m before a set containing a message m', then no other process delivers first a set containing m' and later a set containing m. It has been shown that SCD-broadcast and read/write registers are computationally equivalent, and an algorithm implementing SCD-broadcast is known in the context of asynchronous message passing systems prone to crash failures.

We introduce a Byzantine-tolerant SCD-broadcast algorithm in [23], which we call BSCD-broadcast. Our proposed algorithm assumes an underlying basic Byzantine-tolerant reliable broadcast abstraction. We first introduce an intermediary communication primitive, Byzantine FIFO broadcast (BFIFO-broadcast), which we then use as a primitive in our final BSCD-broadcast algorithm. Unlike the original SCD-broadcast algorithm that is tolerant to up to $t < n/2$ crashing processes, and unlike the underlying Byzantine reliable broadcast primitive that is tolerant to up to $t < n/3$ Byzantine processes, our BSCD-broadcast algorithm is tolerant to up to $t < n/4$ Byzantine processes. As an illustration of the high abstraction power provided by the BSCD-broadcast primitive, we show that it can be used to implement a Byzantine-tolerant read/write snapshot object in an extremely simple way.

6.3.6. PnyxDB: a Lightweight Leaderless Democratic Byzantine Fault Tolerant Replicated Datastore

Participants: Loïck Bonniot, François Taïani.

Byzantine-Fault-Tolerant (BFT) systems are rapidly emerging as a viable technology for production-grade systems, notably in closed consortia deployments for financial and supply-chain applications. Unfortunately, most algorithms proposed so far to coordinate these systems suffer from substantial scalability issues, mainly due to the requirement of a single leader node. We observed that many application workloads offer little concurrency, and proposed PnyxDB, an eventually-consistent BFT replicated datastore that exhibits both high scalability and low latency. Our approach (proposed in [40]) is based on conditional endorsements, that allow nodes to specify the set of transactions that must *not* be committed for the endorsement to be valid.

Additionally, although most of prior art rely on internal voting or quorum mechanisms, these mechanisms are not exposed to applications as first-class primitives. As a result, individual nodes cannot implement application-defined policies without additional effort, costs, and complexity. This is problematic, as application-level voting capabilities are key to a number of emerging decentralized BFT applications involving independent participants who need to balance conflicting goals and shared interests. In addition to its high scalability, PnyxDB supports application-level voting by design. We provided a comparison against BFTS-MaRt and Tendermint, two competitors with different design aims, and demonstrated that our implementation speeds up commit latencies by a factor of 11, remaining below 5 seconds in a worldwide geodistributed deployment of 180 nodes.

PnyxDB's source code is freely available [0]. This work has also been done in collaboration with Christoph Neumann at InterDigital.

6.3.7. Vertex Coloring with Communication Constraints in Synchronous Broadcast Networks

Participants: Hicham Lakhlef, Michel Raynal, François Taïani.

In this work [17], we consider distributed vertex-coloring in broadcast/receive networks suffering from conflicts and collisions. (A collision occurs when, during the same round, messages are sent to the same process by too many neighbors; a conflict occurs when a process and one of its neighbors broadcast during the same round.) More specifically, our work focuses on multi-channel networks, in which a process may either broadcast a message to its neighbors or receive a message from at most γ of them. The work first provides a new upper bound on the corresponding graph coloring problem (known as frugal coloring) in general graphs, proposes an exact bound for the problem in trees, and presents a deterministic, parallel, color-optimal, collision- and conflict-free distributed coloring algorithm for trees, and proves its correctness.

[0] https://github.com/technicolor-research/pnyxdb

6.3.8. *Efficient Randomized Test-and-Set Implementations*

Participant: George Giakkoupis.

In [16], we study randomized test-and-set (TAS) implementations from registers in the asynchronous shared memory model with n processes. We introduce the problem of *group election*, a natural variant of leader election, and propose a framework for the implementation of TAS objects from group election objects. We then present two group election algorithms, each yielding an efficient TAS implementation. The first implementation has expected maxstep complexity $O(\log^* k)$ in the location-oblivious adversary model, and the second has expected maxstep complexity $O(\log \log k)$ against any read/write-oblivious adversary, where $k \leq n$ is the contention. These algorithms improve the previous upper bound by Alistarh and Aspnes (2011) of $O(\log \log n)$ expected maxstep complexity in the oblivious adversary model.

We also propose a modification to a TAS algorithm by Alistarh, Attiya, Gilbert, Giurgiu, and Guerraoui (2010) for the strong adaptive adversary, which improves its space complexity from super-linear to linear, while maintaining its $O(\log n)$ expected maxstep complexity. We then describe how this algorithm can be combined with any randomized TAS algorithm that has expected maxstep complexity $T(n)$ in a weaker adversary model, so that the resulting algorithm has $O(\log n)$ expected maxstep complexity against any strong adaptive adversary and $O(T(n))$ in the weaker adversary model.

Finally, we prove that for any randomized 2-process TAS algorithm, there exists a schedule determined by an oblivious adversary such that with probability at least $1/4^t$ one of the processes needs at least t steps to finish its TAS operation. This complements a lower bound by Attiya and Censor-Hillel (2010) on a similar problem for $n \geq 3$ processes.

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

6.4. Machine Learning and Security

6.4.1. *Adversarial Frontier Stitching for Remote Neural Network Watermarking*

Participant: Erwan Le Merrer.

The state-of-the-art performance of deep learning models comes at a high cost for companies and institutions, due to the tedious data collection and the heavy processing requirements. Recently, Nagai et al. proposed to watermark convolutional neural networks for image classification, by embedding information into their weights. While this is a clear progress toward model protection, this technique solely allows for extracting the watermark from a network that one accesses locally and entirely. Instead, we aim at allowing the extraction of the watermark from a neural network (or any other machine learning model) that is operated remotely, and available through a service API. To this end, we propose in this work [18] to mark the model's action itself, tweaking slightly its decision frontiers so that a set of specific queries convey the desired information. In this work, we formally introduce the problem and propose a novel zero-bit watermarking algorithm that makes use of adversarial model examples. While limiting the loss of performance of the protected model, this algorithm allows subsequent extraction of the watermark using only few queries. We experimented the approach on three neural networks designed for image classification, in the context of the MNIST digit recognition task.

This work was done in collaboration with Gilles Trédan (LAAS/CRNS) and Patrick Pérez (Valéo AI).

6.4.2. *TamperNN: Efficient Tampering Detection of Deployed Neural Nets*

Participant: Erwan Le Merrer.

Neural networks are powering the deployment of embedded devices and Internet of Things. Applications range from personal assistants to critical ones such as self-driving cars. It has been shown recently that models obtained from neural nets can be trojaned ; an attacker can then trigger an arbitrary model behavior facing crafted inputs. This has a critical impact on the security and reliability of those deployed devices. In this work [33], we introduce novel algorithms to detect the tampering with deployed models, classifiers in particular. In the remote interaction setup we consider, the proposed strategy is to identify markers of the model input space that are likely to change class if the model is attacked, allowing a user to detect a possible tampering. This setup makes our proposal compatible with a wide range of scenarios, such as embedded models, or models exposed through prediction APIs. We experiment those tampering detection algorithms on the canonical MNIST dataset, over three different types of neural nets, and facing five different attacks (trojaning, quantization, fine-tuning, compression and watermarking). We then validate over five large models (VGG16, VGG19, ResNet, MobileNet, DenseNet) with a state of the art dataset (VGGFace2), and report results demonstrating the possibility of an efficient detection of model tampering.

This work was done in collaboration with Gilles Trédan (LAAS/CRNS).

6.4.3. MD-GAN: Multi-Discriminator Generative Adversarial Networks for Distributed Datasets

Participant: Erwan Le Merrer.

A recent technical breakthrough in the domain of machine learning is the discovery and the multiple applications of Generative Adversarial Networks (GANs). Those generative models are computationally demanding, as a GAN is composed of two deep neural networks, and because it trains on large datasets. A GAN is generally trained on a single server. In this work, we address the problem of distributing GANsso that they are able to train over datasets that are spread on multiple workers. In this work [31] MD-GAN is exposed as the first solution for this problem: we propose a novel learning procedure for GANs so that they fit this distributed setup. We then compare the performance of MD-GAN to an adapted version of Federated Learning to GANs, using the MNIST and CIFAR10 datasets.MD-GAN exhibits a reduction by a factor of two of the learning complexity on each worker node, while providing better performances than federated learning on both datasets. We finally discuss the practical implications of distributing GANs.

This work was done in collaboration with Bruno Sericola (Inria) and Corentin Hardy (Technicolor).

6.5. Network and Graph Algorithms

6.5.1. Multisource Rumor Spreading with Network Coding

Participants: David Bromberg, Quentin Dufour, Davide Frey.

The last decade has witnessed a rising interest in Gossip protocols in distributed systems. In particular, as soon as there is a need to disseminate events, they become a key functional building block due to their scalability, robustness and fault tolerance under high churn. However, Gossip protocols are known to be bandwidth intensive. A huge amount of algorithms has been studied to limit the number of exchanged messages using different combinations of push/pull approaches. In this work we revisited the state of the art by applying Random Linear Network Coding to further increase performance. In particular, the originality of our approach consists in combining sparse-vector encoding to send our network-coding coefficients and Lamport timestamps to split messages in generations in order to provide efficient gossiping. Our results demonstrate that we are able to drastically reduce bandwidth overhead and dissemination delay compared to the state of the art. We published our results at INFOCOM 2019 [27].

6.5.2. DiagNet: towards a generic, Internet-scale root cause analysis solution

Participants: Loïck Bonniot, François Taïani.

Internet content providers and network operators allocate significant resources to diagnose and troubleshoot problems encountered by end-users, such as service quality of experience degradations. Because the Internet is decentralized, the cause of such problems might lie anywhere between an end-user's device and the service datacenters. Further, the set of possible problems and causes cannot be known in advance, making it impossible to train a classifier with all combinations of faults, causes and locations. We explored how machine learning can be used for Internet-scale root cause analysis using measurements taken from end-user devices: our solution, DiagNet, is able to build generic models that (i) do not make any assumption on the underlying network topology, (ii) do not require to define the full set of possible causes during training, and (iii) can be quickly adapted to diagnose new services.

DiagNet adapts recent image analysis tactics for system and network metrics, collected from a large and dynamic set of landmark servers. In details, it applies non-overlapping convolutions and global pooling to extract generic information about the analyzed network. This genericness allows to build a general model, that can later be generalized to any Internet service with minimal effort. DiagNet leverages backpropagation attention mechanisms to extend the possible root causes to the set of available metrics, making the model fully extensible. We evaluated DiagNet on geodistributed mockup web services and automated users running in 6 AWS regions, and demonstrated promising root cause analysis capabilities. While this initial work is being reviewed, we are deploying DiagNet for real web services and users to evaluate its performance in a more realistic setup.

Christoph Neumann (InterDigital) actively participated in this work.

6.5.3. Application-aware adaptive partitioning for graph processing systems

Participant: Erwan Le Merrer.

Modern online applications value real-time queries over fresh data models. This is the case for graph-based applications, such as social networking or recommender systems,running on front-end servers in production. A core problem in graph processing systems is the efficient partitioning of the input graph over multiple workers. Recent advances over Bulk Synchronous Parallel processing systems (BSP) enabled computations over partitions on those workers, independently of global synchronization supersteps. A good objective partitioning makes the understanding of the load balancing and communication trade-off mandatory for performance improvement. This work [32] addresses this trade-off through the proposal of an optimization problem, that is to be solved continuously to avoid performance degradation over time. Our simulations show that the design of the software module we propose yields significant performance improvements over the BSP processing model.

This work was done in collaboration with Gilles Trédan (LAAS/CRNS).

6.5.4. How to Spread a Rumor: Call Your Neighbors or Take a Walk?

Participant: George Giakkoupis.

In [28], we study the problem of randomized information dissemination in networks. We compare the now standard push-pull protocol, with agent-based alternatives where information is disseminated by a collection of agents performing independent random walks. In the visit-exchange protocol, both nodes and agents store information, and each time an agent visits a node, the two exchange all the information they have. In the meet-exchange protocol, only the agents store information, and exchange their information with each agent they meet.

We consider the broadcast time of a single piece of information in an n-node graph for the above three protocols, assuming a linear number of agents that start from the stationary distribution. We observe that there are graphs on which the agent-based protocols are significantly faster than push-pull, and graphs where the converse is true. We attribute the good performance of agent-based algorithms to their inherently fair bandwidth utilization, and conclude that, in certain settings, agent-based information dissemination, separately or in combination with push-pull, can significantly improve the broadcast time.

The graphs considered above are highly non-regular. Our main technical result is that on any regular graph of at least logarithmic degree, push-pull and visit-exchange have the same asymptotic broadcast time. The proof uses a novel coupling argument which relates the random choices of vertices in push-pull with the random walks in visit-exchange. Further, we show that the broadcast time of meet-exchange is asymptotically at least as large as the other two's on all regular graphs, and strictly larger on some regular graphs.

As far as we know, this is the first systematic and thorough comparison of the running times of these very natural information dissemination protocols.

This work was done in collaboration with Frederik Mallmann-Trenn (MIT) and Hayk Saribekyan (University of Cambridge, UK).

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. CIFRE Technicolor: Distributed troubleshooting of edge-compute functions (2018-2021)

Participants: Loïck Bonniot, François Taïani.

This project seeks to explore how recent generations of end-user gateways (or more generally end-user devices) could implement an edge-compute paradigm powered by user-side micro-services. Our vision is that the devices distributed among the homes of end-users will expose (as a service) their computing power and their ability to quickly deploy compute functions in an execution environment. In order for service and application providers to actually use the system and deploy applications, the system must however ensure an appropriate level of reliability, while simultaneously requiring a very low level of maintenance in order to address the typical size and economics of gateway deployments (at least a few tens of million units). Providing a good level of reliability in such a large system at a reasonable cost is unfortunately difficult. To address this challenge, we aim in this thesis to exploit the *natural distribution* of such large-scale user-side device deployments to quickly pinpoint problems and troubleshoot applications experiencing performance degradations.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. Web of Browser's (Brittany Region and Labex CominLabs 2019-2020)

Participant: François Taïani.

Browsers are de facto the most widely deployed execution environments in the world. Initially simple HTML readers, they now run complex applications interacting with humans and web services. The recent introduction of WebRTC has further extended the capability of browsers by introducing support for browser-to-browser communication. This turns browsers into a decentralized execution environment where interactions between human and web services are enabled without third party.

The Web of browsers is a vision where the web is serverless, ephemeral and massively decentralized. Web where pages are hosted by networks of browsers connected through WebRTC. The objective of the project is to build and experiment the Web of Browsers.

8.2. National Initiatives

8.2.1. ANR Project PAMELA (2016-2020)

Participants: Davide Frey, George Giakkoupis, François Taïani.

PAMELA is a collaborative ANR project involving Inria/IRISA, Inria Lille (MAGNET team), 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 fundamental answers to modern information systems built by interconnecting many personal devices holding private user data in the search of personalized suggestions and recommendations. 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.

8.2.2. ANR Project OBrowser (2016-2020)

Participants: David Bromberg, Davide Frey, François Taïani.

OBrowser is a collaborative ANR project involving Inria, the University of Nantes, the University of South Brittany, and Orange. The project emerges from the vision of designing and deploying distributed applications 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.

8.2.3. ANR Project DESCARTES (2016-2020)

Participants: George Giakkoupis, Michel Raynal, François Taïani.

DESCARTES is a collaborative ANR project involving Inria/IRISA, Labri (U. Bordeaux), IRIF (U. Paris Diderot), Inria Paris (GANG Team), Vérimag (Grenoble), LIF (Marseilles), and LS2N (former LINA, Nantes). The DESCARTES project aims at bridging the lack of a generic theoretical framework in order to unify the large body of fundamental knowledge on distributed computation that has been acquired over the last 40 years. In particular, the project's objective is to develop 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.

8.2.4. Labex CominLab PROFILE (2016-2019)

Participants: David Bromberg, Davide Frey, François Taïani.

The PROFILE (2016-2019) project brings together experts from law, computer science (the Inria teams DIVERSE and ASAP/WIDE, the IRISA team DRUID) 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: first, the project will investigate, design, and introduce a new right of opposition into privacy Law to better regulate profiling and to modify the behavior of commercial companies. Second, the project aims to provide users with the technical means they need to detect stealthy profiling techniques, and to control the extent of the digital traces they routinely produce.

8.3. International Initiatives

8.3.1. 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 is a step in this inescapable direction.

8.4. International Research Visitors

Roberto Rodrigues Filho (Lancaster University, UK), July–September 2019.

Mohamed Lechiakh, (ENSIAS, Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes, Rabat, Morocco), March–May 2019.

Hasnaa Dyani, (ENSIAS, Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes, Rabat, Morocco), April–June 2019.

Chaimaa Tarzi, (ENSIAS, Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes, Rabat, Morocco), April–June 2019.

Arsany Guirguis, (EPFL, Lausanne, Switzerland), July–September 2019.

Marcus Kaboret, (Laboratoire de Mathématiques et Informatique, Joseph Ki-Zerbo University, Ouagadougou, Burkina Faso), September–October 2019.

Hayk Saribekyan (University of Cambridge, UK), 2-–12 April 2019.

Giorgi Nadiradze (IST Austria), 20–24 May 2019.

Emanuele Natale (CNRS, Sophia-Antipolis), 13–17 March 2019.

8.4.1. Visits to International Teams

Adrien Luxey visited Paulo Ferreira, University of Oslo, Norway, from the 1st May to the 30th of June 2019.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Member of the Organizing Committees

François Taïani and Erwan Le Merrer served as Co-Organizer of the 9th Inria/Interdigtal Workshop On Systems (WOS9), Rennes, France, December 2019.

François Taïani served as Posters and Demos Co-Chair of the 38th IEEE International Symposium on Reliable Distributed Systems (SRDS 2019), Lyon, France, October 2019.

Erwan Le Merrer served as Co-Organizer of the "Atelier sur les algorithmes en boite-noire", Lyon, France, December 2019.

9.1.2. Scientific Events Selection

9.1.2.1. Chair of Conference Program Committees

François Taïani served as PC Co-Chair of the 3^{rd} Workshop on Scalable and Resilient Infrastructures for Distributed Ledgers (SERIAL'19), collocated with Middleware 2019, Davis, CA, USA, December 2019.

François Taïani served as Co-Chair of the Shadow PC of the European Conference on Computer Systems (EuroSys 2019), Dresden, Germany, March 2019.

Erwan Le Merrer served as Co-Chair of the "Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications" (Algotel 2019), France, June 2019.

9.1.2.2. Member of the Conference Program Committees

François Taïani served on the PC of the Conférence d'informatique en Parallélisme, Architecture et Système (COMPAS2019), Anglet, France, June 2019.

François Taïani served on the PC of the 18th Workshop on Adaptive and Reflexive Middleware (ARM'19@MW), co-located with Middleware 2019, Davis, CA, USA, December 2019.

François Taïani served on the PC of the 49th IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2019), Portaland, OR, USA, June 2019.

François Taïani served on the PC of the International Conference on Blockchain Economics, Security and Protocols (Tokenomics 2019), Paris, France, May 2019.

Davide Frey served on the PC of the 5th Workshop on Planetary-Scale Distributed Systems (W-PSDS), co-located with SRDS 2019, Lyon, France, October 2019.

Davide Frey served on the PC of the 13th ACM International Conference on Distributed and Event-Based Systems (DEBS), Darmstadt, Germany, June 2019.

Davide Frey served on the PC of the 19th International Conference on Distributed Applications and Interoperable Systems (DAIS), Copenhagen, Denmark, June 2019.

George Giakkoupis served on the PC of the 31st ACM Symposium on Parallelism in Algorithms and Architectures (SPAA), Phoenix, AZ, USA, June 22–24, 2019.

George Giakkoupis served on the PC of the 33rd International Symposium on Distributed Computing (DISC), Budapest, Hungary, Oct. 14–18, 2019.

9.1.3. Journal

9.1.3.1. Reviewer - Reviewing Activities

François Taïani was a reviewer for IEEE Transactions on Knowledge and Data Engineering (TKDE).

Davide Frey was a reviewer for the International Journal of Computer and Telecommunications Networking (COMNET).

Davide Frey was a reviewer for Peer-to-Peer Networking and Applications (PPNA).

Davide Frey was a reviewer for the ARIMA Journal.

George Giakkoupis was a reviewer for Journal of the ACM (JACM).

George Giakkoupis was a reviewer for Journal of Parallel and Distributed Computing (JPDC).

George Giakkoupis was a reviewer for ACM Transactions on Parallel Computing (TOPC).

George Giakkoupis was a reviewer for Distributed Computing (DIST).

9.1.4. Collaborative Projects

Davide Frey was a reviewer for the Agence Nationale de Recherche (ANR).

9.1.5. Invited Talks

François Taïani. Pleiades: Distributed Structural Invariants at Scale. 4th GDR RSD and ASF Winter School on Distributed Systems and Networks 2019, Pleynet, Sept Laux, France, 6 February 2019.

François Taïani. Pleiades: Distributed Structural Invariants at Scale. Lancaster University, Lancaster, UK, 23 May 2019.

Erwan Le Merrer. Tweaking neural models: watermarking and tamperproofing them. EPFL-Inria workshop, EPFL, Lausanne, 30 January 2019.

François Taïani. How to Be a Terrible Higher Education Teacher in 7 Easy Steps. IRISA Annual D1 Departmental Seminar, IRISA, Rennes, 29 November 2019.

9.1.6. Leadership within the Scientific Community

Michel Raynal has been Adjunct Professor at the Polytechnic University of Hong Kong since 2013.

Michel Raynal has been a member of the Executive board of the SIF (Société d'Informatique Française) since 2013.

Michel Raynal is European representative at the IEEE Technical Committee on distributed computing.

François Taïani has been a member of the Gilles Kahn PhD Award in Computer Science, from Société Informatique de France (SIF) since 2018.

9.1.7. Research Administration

Davide Frey is Correspondant Scientifique Europe at the DPEI for Inria Rennes.

Davide Frey is an associate member of the COST-GTRI of Inria.

François Taïani is a Career Advice Person, (*Référent conseil-parcours professionel chercheurs*) for IRISA/Inria Rennes Bretagne Atlantique since 2019.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Master: Florestan De Moor, Programmation Dirigée par la Syntaxe, 22h, M1-SIF, Université de Rennes 1, France

Engineering School: François Taïani, Synchronization and Parallel Programming, 62h, 2nd year of Engineering School (M1), ESIR / Univ. Rennes I, France.

Engineering School: François Taïani, Distributed Systems, 24h, 3rd year of Engineering School (M2), ESIR / Univ. Rennes I, France.

Engineering School: François Taïani, Introduction to Operating Systems, 24h, 1st year of Engineering School (L3), ESIR / Univ. Rennes I, France.

Bacherlor: François Taïani, Distributed Algorithms, 12h, L3 Parcours SI, ENS Rennes, France.

Master: Davide Frey, 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.

Master: Davide Frey, Cloud Computing, 6 hours, M2-MIAGE, Univ. Rennes I, France.

Master: Davide Frey, Distributed Systems, 12 hours, ENSAI, France.

Master: Davide Frey, Apprentice Tutoring, 8 ETD hours, M2 Alternance Univ. Rennes I, France.

Master: Quentin Dufour and Davide Frey, 12 ETD hours, M1 ISTIC, Univ. Rennes I, France.

Master / PhD: Davide Frey, Distributed Computing and Blockchain, 15 hours, UM6P, Morocco.

Master: Erwan Le Merrer, Projet , 36 ETD hours, M1 ISTIC, Univ. Rennes I, France.

Master: Erwan Le Merrer, Network Science, 12 hours, M2 ESIR, Univ. Rennes I, France.

Master: Quentin Dufour and Louison Gitzinger, Cloud, 36 hours, M2 ESIR, Univ. Rennes I, France.

Master: Quentin Dufour, TLC, 18 hours, M2 ISTIC, Univ. Rennes I, France

Master: George Giakkoupis, Systèmes Répartis, 9 hours, ENSAI Rennes, France.

9.2.2. Supervision

PhD: Adrien Luxey, "E-squads: A novel paradigm to build privacy-preserving ubiquitous applications" [13], University of Rennes 1, 29th November 2019, David Bromberg.

PhD in progress: Quentin Dufour, BBDA - Browser Based Data Analytics, January 2018, David Bromberg and Davide Frey.

PhD in progress: Amaury Bouchra Pilet, Robust and Lightweight Overlay Management for Decentralized Learning, University of Rennes 1, September 2018, David Bromberg and Davide Frey.

PhD in progress: Loïck Bonniot, Distributed Troubleshooting of Edge-Compute Functions, University of Rennes 1, François Taïani and Christoph Neumann (Interdigital, CIFRE).

PhD in progress: Alejandro Gomez Boix, Distributed counter-measure against browser fingerprinting, Inria, Davide Frey and David Bromberg.

PhD in progress: Alex Auvolat, Towards probabilistic decentralized systematic design for large-scale privacy-preserving collaborative systems, University of Rennes 1, François Taïani and David Bromberg.

PhD in progress: Hayk Saribekyan, Randomized Algorithms for Distributed Information Dissemination, University of Cambridge, UK, George Giakkoupis and Thomas Sauerwald (U. Cambridge).

HDR: Davide Frey, Epidemic Protocols: From Large Scale to Big Data, University of Rennes 1, 11 Juin 2019 [12].

9.2.3. Juries

François Taïani was an examiner for Cédric Maigrot's PhD thesis: Détection de fausses informations dans les réseaux sociaux, Université de Rennes 1 (France), 1st April 2019.

François Taïani was a reviewer for João Paulo de Araujo's PhD thesis: A Communication-Efficient Causal Broadcast Publish/Subscribe System, Sorbonne Université, LIP6 (France), 5th April 2019.

François Taïani was an external examiner (reviewer) for Assylbek Sagitzhanuly Jumagaliyev's PhD thesis: A Modeling Language for Multi-tenant Data Architecture Evolution in Cloud Applications, Lancaster University (UK), 23rd May 2019.

François Taïani was an examiner for Gilles Tredan's HDR: Capturing Binary Graphs, Université de Toulouse 3 (France), 21st June 2019.

François Taïani was an examiner for Nathanaël Cheriere's PhD thesis: Towards Malleable Distributed Storage Systems: from Models to Practice, ENS Rennes, 5th November 2019.

François Taïani was an examiner for The Anh Pham's PhD thesis: Efficient state-space exploration for asynchronous distributed programs, ENS Rennes, 6th December 2019.

François Taïani was an examiner for Oscar Luis Vera Perez's PhD thesis: Dynamic program analysis for suggesting test improvements to developers, University of Rennes 1, 17th December 2019.

George Giakkoupis was an external reviewer for Suman Sourav's PhD thesis: Latency, Conductance, and the Role of Connectivity in Graph Problems, National University of Singapore, 11 November 2019.

9.3. Popularization

Algorithm Watch wrote an article about a recent work involving Erwan Le Merrer, concernant remote transparency. The article is called "Explainable AI doesn't work for online services – now there's proof". Avaiable at https://algorithmwatch.org/en/story/explainable-ai-doesnt-work-for-online-services-now-theres-proof/.

10. Bibliography

Major publications by the team in recent years

[1] 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, ACM, January 2018, p. 1715-1733, https://hal.inria.fr/hal-01635757

[2] S. BOUGET, Y.-D. BROMBERG, A. LUXEY, F. TAÏANI.*Pleiades: Distributed Structural Invariants at Scale*, in "DSN 2018 - IEEE/IFIP International Conference on Dependable Systems and Networks", Luxembourg, Luxembourg, IEEE, June 2018, p. 542-553 [*DOI :* 10.1109/DSN.2018.00062], https://hal.archives-ouvertes.fr/hal-01803881

[3] Z. BOUZID, M. RAYNAL, P. SUTRA.*Anonymous obstruction-free (n, k)-set agreement with n-k+1 atomic read/write registers*, in "Distributed Computing", April 2018, vol. 31, n^{o} 2, p. 99-117, https://hal.inria.fr/hal-01952626

[4] Y.-D. BROMBERG, Q. DUFOUR, D. FREY.*Multisource Rumor Spreading with Network Coding*, in "INFOCOM 2019 - IEEE International Conference on Computer Communications", Paris, France, IEEE, April 2019, p. 1-10, https://hal.inria.fr/hal-01946632

[5] Y.-D. BROMBERG, A. LUXEY, F. TAÏANI.*CASCADE: Reliable Distributed Session Handoff for Continuous Interaction across Devices*, in "ICDCS 2018 - 38th IEEE International Conference on Distributed Computing Systems", Vienna, Austria, IEEE, July 2018, p. 244-254 [*DOI :* 10.1109/ICDCS.2018.00033], https://hal.inria.fr/hal-01797548

[6] 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^o 2, p. 19:1 - 19:26 [*DOI :* 10.1145/3086465], https://hal.inria.fr/hal-01635743

[7] F. CHIERICHETTI, G. GIAKKOUPIS, S. LATTANZI, A. PANCONESI.*Rumor Spreading and Conductance*, in "Journal of the ACM (JACM)", August 2018, vol. 65, n^o 4, p. 17:1-17:21 [*DOI :* 10.1145/3173043], https://hal.inria.fr/hal-01942162

[8] 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

[9] H. LAKHLEF, M. RAYNAL, F. TAÏANI.*Vertex Coloring with Communication Constraints in Synchronous Broadcast Networks*, in "IEEE Transactions on Parallel and Distributed Systems", July 2019, vol. 30, n^o 7, p. 1672-1686 [*DOI :* 10.1109/TPDS.2018.2889688], https://hal.inria.fr/hal-02376726

[10] A. LUXEY, Y.-D. BROMBERG, F. M. COSTA, V. LIMA, R. DA ROCHA, F. TAÏANI.*Sprinkler: A probabilistic dissemination protocol to provide fluid user interaction in multi-device ecosystems*, in "PerCom 2018 - IEEE International Conference on Pervasive Computing and Communications", Athens, Greece, IEEE, March 2018, p. 1-10 [*DOI :* 10.1109/PERCOM.2018.8444577], https://hal.inria.fr/hal-01704172

[11] 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

Publications of the year

Doctoral Dissertations and Habilitation Theses

[12] D. FREY.*Epidemic Protocols: From Large Scale to Big Data*, Université De Rennes 1, June 2019, Habilitation à diriger des recherches, https://hal.inria.fr/tel-02375909

[13] A. LUXEY.*E-squads: A novel paradigm to build privacy-preserving ubiquitous applications*, Université de Rennes, November 2019, https://hal.inria.fr/tel-02389297

Articles in International Peer-Reviewed Journal

[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 2019, p. 344–365 [*DOI :* 10.1007/S00224-017-9772-Y], https://hal.archives-ouvertes.fr/hal-01672723

[15] D. FREY, M. X. MAKKES, P.-L. ROMAN, F. TAÏANI, S. VOULGARIS.*Dietcoin: Hardening Bitcoin Transaction Verification Process For Mobile Devices*, in "Proceedings of the VLDB Endowment (PVLDB)", August 2019, vol. 12, n^o 12, p. 1946-1949 [*DOI :* 10.14778/3352063.3352106], https://hal.inria.fr/hal-02315154

[16] G. GIAKKOUPIS, P. WOELFEL.*Efficient Randomized Test-And-Set Implementations*, in "Distributed Computing", 2019, p. 565–586, forthcoming [*DOI :* 10.1007/S00446-019-00349-Z], https://hal.inria.fr/hal-02012672

[17] H. LAKHLEF, M. RAYNAL, F. TAÏANI.*Vertex Coloring with Communication Constraints in Synchronous Broadcast Networks*, in "IEEE Transactions on Parallel and Distributed Systems", July 2019, vol. 30, n^o 7, p. 1672-1686 [*DOI :* 10.1109/TPDS.2018.2889688], https://hal.inria.fr/hal-02376726

[18] E. LE MERRER, P. PÉREZ, G. TRÉDAN.*Adversarial frontier stitching for remote neural network watermarking*, in "Neural Computing and Applications", 2019, p. 1-12 [*DOI :* 10.1007/S00521-019-04434-Z], https://hal.archives-ouvertes.fr/hal-02264449

[19] A. MOSTEFAOUI, M. PERRIN, M. RAYNAL, J. CAO.*Crash-Tolerant Causal Broadcast in O(n) Messages*, in "Information Processing Letters", November 2019, vol. 151, p. 1-9 [*DOI :* 10.1016/J.IPL.2019.105837], https://hal.archives-ouvertes.fr/hal-02279523

[20] A. MOSTEFAOUI, M. RAYNAL, M. ROY.*Time-Efficient Read/Write Register in Crash-prone Asynchronous Message-Passing Systems*, in "Computing", January 2019, vol. 101, n^o 1, p. 3-17 [*DOI :* 10.1007/s00607-018-0615-8], https://hal.laas.fr/hal-01784210

[21] J. OLIVARES, A.-M. KERMARREC, N. CHILUKA.*The out-of-core KNN awakens: the light side of computation force on large datasets*, in "Computing", 2019, https://hal.inria.fr/hal-02387161

International Conferences with Proceedings

[22] Z. AGHAZADEH, D. IMBS, M. RAYNAL, G. TAUBENFELD, P. WOELFEL.*Optimal Memory-Anonymous Symmetric Deadlock-Free Mutual Exclusion*, in "PODC 2019 - ACM Symposium on Principles of Distributed Computing", Toronto, Canada, ACM, July 2019, p. 157-166 [*DOI :* 10.1145/3293611.3331594], https://hal.archives-ouvertes.fr/hal-02394246

[23] A. AUVOLAT, M. RAYNAL, F. TAÏANI.*Byzantine-Tolerant Set-Constrained Delivery Broadcast*, in "OPODIS 2019 - International Conference on Principles of Distributed Systems", Neuchâtel, Switzerland, ACM, December 2019, p. 1-23 [*DOI :* 10.4230/LIPICS.OPODIS.2019.16], https://hal.inria.fr/hal-02376673

[24] A. AUVOLAT, F. TAÏANI.*Merkle Search Trees: Efficient State-Based CRDTs in Open Networks*, in "SRDS 2019 - 38th IEEE International Symposium on Reliable Distributed Systems", Lyon, France, IEEE, October 2019, p. 1-10 [*DOI :* 10.1109/SRDS.2019.00032], https://hal.inria.fr/hal-02303490

[25] O. BALMAU, R. GUERRAOUI, A.-M. KERMARREC, A. MAURER, M. PAVLOVIC, W. ZWAENEPOEL.*The Fake News Vaccine - A Content-Agnostic System for Preventing Fake News from Becoming Viral*, in "NETYS 2019 - 7th International Conference on NETworked sYStems", Marrakech, Morocco, Springer, June 2019, p. 347-364 [*DOI :* 10.1007/978-3-030-31277-0_23], https://hal.inria.fr/hal-02387182

[26] A. BOUCHRA PILET, D. FREY, F. TAÏANI.*Robust Privacy-Preserving Gossip Averaging*, in "SSS 2019 - 21st International Symposium on Stabilization, Safety, and Security of Distributed Systems", Pisa, Italy,

Springer, November 2019, p. 38-52 [*DOI :* 10.1007/978-3-030-34992-9_4], https://hal.archives-ouvertes.fr/hal-02373353

[27] Y.-D. BROMBERG, Q. DUFOUR, D. FREY.*Multisource Rumor Spreading with Network Coding*, in "INFOCOM 2019 - IEEE International Conference on Computer Communications", Paris, France, IEEE, April 2019, p. 1-10, https://hal.inria.fr/hal-01946632

[28] G. GIAKKOUPIS, F. MALLMANN-TRENN, H. SARIBEKYAN.*How to Spread a Rumor: Call Your Neighbors or Take a Walk?*, in "PODC 2019 - ACM Symposium on Principles of Distributed Computing", Toronto ON, Canada, ACM Press, July 2019, p. 24-33 [*DOI :* 10.1145/3293611.3331622], https://hal.inria.fr/hal-02388328

[29] R. GUERRAOUI, A.-M. KERMARREC, O. RUAS, F. TAÏANI.*Fingerprinting Big Data: The Case of KNN Graph Construction*, in "ICDE 2019 - 35th IEEE International Conference on Data Engineering", Macao, China, IEEE, April 2019, p. 1738-1741 [*DOI :* 10.1109/ICDE.2019.00186], https://hal.inria.fr/hal-02357950

[30] A. GÓMEZ-BOIX, D. FREY, Y.-D. BROMBERG, B. BAUDRY.*A Collaborative Strategy for mitigating Tracking through Browser Fingerprinting*, in "MTD 2019 - 6th ACM Workshop on Moving Target Defense", London, United Kingdom, November 2019, p. 1-12 [*DOI :* 10.1145/3338468.3356828], https://hal.inria.fr/hal-02282591

[31] C. HARDY, E. LE MERRER, B. SERICOLA.*MD-GAN: Multi-Discriminator Generative Adversarial Networks for Distributed Datasets*, in "IPDPS 2019 - 33rd IEEE International Parallel and Distributed Processing Syposium", Rio de Janeiro, Brazil, IEEE, May 2019, p. 1-12, https://hal.inria.fr/hal-01946665

[32] E. LE MERRER, G. TRÉDAN.*Application-aware adaptive partitioning for graph processing systems*, in "MASCOTS 2019 - 27th IEEE International Symposium on the Modeling, Analysis, and Simulation of Computer and Telecommunication Systems", Rennes, France, IEEE, October 2019, p. 235-240 [*DOI :* 10.1109/MASCOTS.2019.00033], https://hal.archives-ouvertes.fr/hal-02193594

[33] E. LE MERRER, G. TRÉDAN.*TamperNN: Efficient Tampering Detection of Deployed Neural Nets*, in "ISSRE 2019 - IEEE 30th International Symposium on Software Reliability Engineering", Berlin, Germany, October 2019, p. 1-11, https://hal.archives-ouvertes.fr/hal-02268136

[34] M. RAYNAL, J. CAO.*One for All and All for One: Scalable Consensus in a Hybrid Communication Model*, in "ICDCS'19 - 39th IEEE International Conference on Distributed Computing Systems", Dallas, United States, IEEE, July 2019, p. 464-471 [*DOI :* 10.1109/ICDCS.2019.00053], https://hal.archives-ouvertes.fr/hal-02394259

Conferences without Proceedings

[35] C. DELPORTE-GALLET, H. FAUCONNIER, M. RAYNAL.*On the Weakest Failure Detector for Read/Write-Based Mutual Exclusion*, in "AINA 2019 - 33rd International Conference on Advanced Information Networking and Applications", Matsue, Japan, March 2020, https://hal.archives-ouvertes.fr/hal-02433606

[36] A. DURAND, M. RAYNAL, G. TAUBENFELD.*Pannes de processus liées à la contention*, in "ALGOTEL 2019 - 21èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Saint Laurent de la Cabrerisse, France, June 2019, p. 1-4, https://hal.archives-ouvertes.fr/hal-02118917

[37] E. GODARD, D. IMBS, M. RAYNAL, G. TAUBENFELD.*Anonymous Read/Write Memory: Leader Election and De-anonymization*, in "SIROCCO'19 - 26th International Colloquium on Structural Information and Communication Complexity", L'Aquila, Italy, July 2019, p. 246-261 [*DOI :* 10.1007/978-3-030-24922-9_17], https://hal.archives-ouvertes.fr/hal-02445121

[38] E. GODARD, D. IMBS, M. RAYNAL, G. TAUBENFELD.*Mutex-Based De-anonymization of an Anonymous Read/Write Memory*, in "NETYS 2019 - 7th International Conference on Networked Systems", Marrakech, Morocco, June 2019, p. 311-326 [*DOI :* 10.1007/978-3-030-31277-0_21], https://hal.archives-ouvertes.fr/hal-02445119

[39] R. GUERRAOUI, E. LE MERRER, R. PATRA, J.-R. VIGOUROUX.*Unified and Scalable Incremental Recommenders with Consumed Item Packs*, in "EURO-PAR 2019 - European Conference on Parallel Processing", Gottingen, Germany, Springer, August 2019, p. 227-240 [*DOI :* 10.1007/978-3-030-29400-7_17], https://hal.archives-ouvertes.fr/hal-02153388

Other Publications

[40] L. BONNIOT, C. NEUMANN, F. TAÏANI.*PnyxDB: a Lightweight Leaderless Democratic Byzantine Fault Tolerant Replicated Datastore*, November 2019, https://arxiv.org/abs/1911.03291 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-02355778

[41] A. BOUCHRA PILET, D. FREY, F. TAÏANI.*Simple, Efficient and Convenient Decentralized Multi-Task Learning for Neural Networks*, November 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02373338

[42] *Best Paper*
F. DE MOOR.*A Biclustering Approach to Recommender Systems*, University of Rennes 1, June 2019, p. 1-46, https://hal.inria.fr/hal-02369708.

References in notes

[43] Y. AFEK, E. GAFNI.*Asynchrony from synchrony*, in "ICDCN", 2013, p. 225–239

[44] A. AHMED, E. AHMED.*A survey on mobile edge computing*, in "2016 10th International Conference on Intelligent Systems and Control (ISCO)", Jan 2016, p. 1-8, http://dx.doi.org/10.1109/ISCO.2016.7727082

[45] T. ALLARD, D. FREY, G. GIAKKOUPIS, J. LEPILLER.*Lightweight Privacy-Preserving Averaging for the Internet of Things*, in "M4IOT 2016 - 3rd Workshop on Middleware for Context-Aware Applications in the IoT", Trento, Italy, ACM, December 2016, p. 19 - 22 [*DOI :* 10.1145/3008631.3008635], https://hal.inria.fr/hal-01421986

[46] E. ANSHELEVICH, D. CHAKRABARTY, A. HATE, C. SWAMY.*Approximability of the Firefighter Problem: Computing Cuts over Time*, in "Algorithmica", 2012, vol. 62, n^{o} 1-2, p. 520–536

[47] D. BERNSTEIN.*Containers and Cloud: From LXC to Docker to Kubernetes*, in "IEEE Cloud Computing", Sept 2014, vol. 1, n^{o} 3, p. 81-84, http://dx.doi.org/10.1109/MCC.2014.51

[48] M. BERTIER, D. FREY, R. GUERRAOUI, A.-M. KERMARREC, V. LEROY.*The Gossple Anonymous Social Network*, in "ACM/IFIP/USENIX 11th International Middleware Conference (MIDDLEWARE)", Bangalore, India, I. GUPTA, C. MASCOLO (editors), Middleware 2010, Springer, November 2010, vol. LNCS-6452, p. 191-211 [*DOI :* 10.1007/978-3-642-16955-7_10], https://hal.inria.fr/inria-00515693

[49] F. BONOMI.*Connected vehicles, the internet of things, and fog computing. VANET 2011, 2011*, 2011, Keynote speech at VANET

[50] F. BONOMI, R. MILITO, J. ZHU, S. ADDEPALLI.*Fog Computing and Its Role in the Internet of Things*, in "1 s t MCC Workshop on Mobile Cloud Computing", 2012, http://doi.acm.org/10.1145/2342509.2342513

[51] A. BOUTET, D. FREY, R. GUERRAOUI, A. JÉGOU, A.-M. KERMARREC.*Privacy-Preserving Distributed Collaborative Filtering*, in "Computing", August 2016, vol. 98, n^{o} 8, https://hal.inria.fr/hal-01251314

[52] A. BOUTET, D. FREY, R. GUERRAOUI, A.-M. KERMARREC, R. PATRA.*HyRec: Leveraging Browsers for Scalable Recommenders*, in "Middleware 2014", Bordeaux, France, December 2014 [*DOI :* 10.1145/2663165.2663315], https://hal.inria.fr/hal-01080016

[53] A. BOUTET, D. FREY, R. GUERRAOUI, A.-M. KERMARREC, A. RAULT, F. TAÏANI, J. WANG.*Hide & Share: Landmark-based Similarity for Private KNN Computation*, in "45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)", Rio de Janeiro, Brazil, June 2015, p. 263-274 [*DOI :* 10.1109/DSN.2015.60], https://hal.archives-ouvertes.fr/hal-01171492

[54] A. BOUTET, D. FREY, A. JÉGOU, A.-M. KERMARREC, H. RIBEIRO.*FreeRec: an Anonymous and Distributed Personalization Architecture*, in "Computing", December 2013, https://hal.inria.fr/hal-00909127

[55] B. COHEN.*Incentives Build Robustness in BitTorrent*, 2003, http://citeseer.ist.psu.edu/cohen03incentives.html

[56] C. DELPORTE-GALLET, H. FAUCONNIER, R. GUERRAOUI, A. TIELMANN.*The disagreement power of an adversary*, in "Distributed Computing", 2011, vol. 24, n^{o} 3-4, p. 137–147

[57] A. J. DEMERS, D. H. GREENE, C. HAUSER, W. IRISH, J. LARSON, S. SHENKER, H. E. STURGIS, D. C. SWINEHART, D. B. TERRY.*Epidemic Algorithms for Replicated Database Maintenance*, in "PODC", 1987, p. 1–12

[58] D. FREY, R. GUERRAOUI, A.-M. KERMARREC, M. MONOD, K. BORIS, M. MARTIN, V. QUÉMA.*Heterogeneous Gossip*, in "Middleware 2009", Urbana-Champaign, IL, United States, December 2009, https://hal.inria.fr/inria-00436125

[59] W. M. GOLAB, V. HADZILACOS, D. HENDLER, P. WOELFEL.*RMR-efficient implementations of comparison primitives using read and write operations*, in "Distributed Computing", 2012, vol. 25, n^{o} 2, p. 109-162

[60] R. GUERRAOUI, K. HUGUENIN, A.-M. KERMARREC, M. MONOD, S. PRUSTY.*LiFTinG: Lightweight Freerider-Tracking Protocol in Gossip*, in "11th ACM/IFIP/USENIX International Middleware Conference (MIDDLEWARE)", Bangalore, India, November 2010 [*DOI :* 10.1007/978-3-642-16955-7_16], https://hal.inria.fr/inria-00505268

[61] R. A. HOLLEY, T. M. LIGGETT.*Ergodic Theorems for Weakly Interacting Infinite Systems and the Voter Model*, in "The Annals of Probability", 1975, vol. 3, n^{o} 4, p. 643-663

[62] D. IMBS, M. RAYNAL.*A liveness condition for concurrent objects: x-wait-freedom*, in "Concurrency and Computation: Practice and experience", 2011, vol. 23, n^{o} 17, p. 2154–2166

[63] F. JUNQUEIRA, K. MARZULLO.*A framework for the design of dependent-failure algorithms*, in "Concurrency and Computation: Practice and Experience", 2007, vol. 19, n^{o} 17, p. 2255–2269

[64] D. KEMPE, J. M. KLEINBERG, É. TARDOS.*Maximizing the spread of influence through a social network*, in "KDD", 2003, p. 137–146

[65] D. KEMPE, J. M. KLEINBERG, É. TARDOS.*Influential Nodes in a Diffusion Model for Social Networks*, in "ICALP", 2005, p. 1127–1138

[66] D. KEMPE, J. M. KLEINBERG, É. TARDOS.*Maximizing the Spread of Influence through a Social Network*, in "Theory of Computing", 2015, vol. 11, p. 105–147

[67] P. KUZNETSOV.*Understanding non-uniform failure models*, in "Bulletin of the EATCS", 2012, n^{o} 106, p. 53–77

[68] E. LIEBERMAN, C. HAUERT, M. A. NOWAK.*Evolutionary dynamics on graphs*, in "Nature", 2005, vol. 433, n^{o} 7023, p. 312–316

[69] M. RAYNAL, J. STAINER.*Synchrony weakened by message adversaries vs asynchrony restricted by failure detectors*, in "PODC", Montréal, Canada, Proceedings of the 2013 ACM symposium on Principles of distributed computing, ACM, July 2013, p. 166-175 [*DOI :* 10.1145/2484239.2484249], https://hal.inria.fr/hal-00920734

[70] N. SANTORO, P. WIDMAYER.*Time is not a healer*, in "Annual Symposium on Theoretical Aspects of Computer Science", Springer, 1989, p. 304–313

[71] A. VERMA, L. PEDROSA, M. KORUPOLU, D. OPPENHEIMER, E. TUNE, J. WILKES.*Large-scale cluster management at Google with Borg*, in "Tenth European Conference on Computer Systems (Eurosys 2015)", ACM, 2015, 18

[72] L. ZHANG, F. ZHOU, A. MISLOVE, R. SUNDARAM.*Maygh: Building a CDN from Client Web Browsers*, in "8th ACM European Conference on Computer Systems", New York, NY, USA, EuroSys '13, ACM, 2013, p. 281–294, http://doi.acm.org/10.1145/2465351.2465379