
RESEARCH CENTER
**Rennes - Bretagne-Atlantique**

FIELD

Activity Report 2014

# Section New Results

Edition: 2015-03-24

**ALF Project-Team**

# 6. New Results

## 6.1. Highlights of the Year

André Seznec and Pierre Michaud won the 4th Championship Branch Prediction in all the 3 categories, 4KB, 32 KB and unlimited storage predictors [23], [33], thus confirming the past championships in 2011, 2006 and 2004.

## 6.2. Processor Architecture

**Participants:** Pierre Michaud, Bharath Narasimha Swamy, Sylvain Collange, Erven Rohou, André Seznec, Arthur Perais, Surya Khizakanchery Natarajan, Sajith Kalathingal, Tao Sun, Andrea Mondelli, Aswinkumar Sridharan, Alain Ketterlin.

Processor, cache, locality, memory hierarchy, branch prediction, multicore, power, temperature

Multicore processors have now become mainstream for both general-purpose and embedded computing. Instead of working on improving the architecture of the next generation multicore, with the DAL project, we deliberately anticipate the next few generations of multicores. While multicores featuring 1000s of cores might become feasible around 2020, there are strong indications that sequential programming style will continue to be dominant. Even future mainstream parallel applications will exhibit large sequential sections. Amdahl's law indicates that high performance on these sequential sections is needed to enable overall high performance on the whole application. On many (most) applications, the effective performance of future computer systems using a 1000-core processor chip will significantly depend on their performance on both sequential code sections and single threads.

We envision that, around 2020, the processor chips will feature a few complex cores and many (maybe 1000's) simpler, more silicon and power effective cores.

In the DAL research project, http://www.irisa.fr/alf/dal, we explore the microarchitecture techniques that will be needed to enable high performance on such heterogeneous processor chips. Very high performance will be required on both sequential sections, -legacy sequential codes, sequential sections of parallel applications-, and critical threads on parallel applications, -e.g. the main thread controlling the application. Our research focuses essentially on enhancing single process performance.

### *6.2.1. Microarchitecture*

#### *6.2.1.1. Branch prediction*

**Participants:** André Seznec, Pierre Michaud.

We submitted 3 predictors to the 4th Championship Branch Prediction that took place along with the ISCA 2014 conference [33], [22], [23]. Our predictors combine some branch prediction techniques that we introduced in our previous works, in particular TAGE [10] and GEHL [12]. The predictor we submitted to the 4KB and 32KB tracks was ranked first [33] in both tracks. The 3 predictors we submitted to the unlimited-size track took the first three ranks. We have established a new reference point for branch predictability limits [23].

The 12 competing predictors were mostly using already published branch prediction techniques. The main learning from this year's contest is that choosing the right combination of techniques for the given constraints is at least as important as trying to specialize branch predictors for certain branch behaviors.

#### *6.2.1.2. Revisiting Value Prediction*

**Participants:** Arthur Perais, André Seznec.

Value prediction was proposed in the mid 90's to enhance the performance of high-end microprocessors. The research on Value Prediction techniques almost vanished in the early 2000's as it was more effective to increase the number of cores than to dedicate some silicon area to Value Prediction. However high end processor chips currently feature 8-16 high-end cores and the technology will allow to implement 50-100 of such cores on a single die in a foreseeable future. Amdahl's law suggests that the performance of most workloads will not scale to that level. Therefore, dedicating more silicon area to value prediction in high-end cores might be considered as worthwhile for future multicores.

First, we introduce a new value predictor VTAGE harnessing the global branch history [29]. VTAGE directly inherits the structure of the indirect jump predictor ITTAGE [10]. VTAGE is able to predict with a very high accuracy many values that were not correctly predicted by previously proposed predictors, such as the FCM predictor and the stride predictor. Three sources of information can be harnessed by these predictors: the global branch history, the differences of successive values and the local history of values. Moreover, VTAGE does not suffer from short critical prediction loops and can seamlessly handle back-to-back predictions, contrarily to previously proposed, hard to implement FCM predictors.

Second, we show that all predictors are amenable to very high accuracy at the cost of some loss on prediction coverage [29]. This greatly diminishes the number of value mispredictions and allows to delay validation until commit-time. As such, no complexity is added in the out-of-order engine because of VP (save for ports on the register file) and pipeline squashing at commit-time can be used to recover. This is crucial as adding *selective replay* in the OoO core would tremendously increase complexity.

Third, we leverage the possibility of validating predictions at commit to introduce a new microarchitecture, EOLE [28]. EOLE features *Early Execution* to execute simple instructions whose operands are ready in parallel with Rename and *Late Execution* to execute simple predicted instructions and high confidence branches just before Commit. EOLE depends on Value Prediction to provide operands for *Early Execution* and predicted instructions for *Late Execution*. However, Value Prediction requires EOLE to become truly practical. That is, EOLE allows to reduce the out-of-order issue-width by 33% without impeding performance. As such, the number of ports on the register file diminishes. Furthermore, optimizations of the register file such as *banking* further reduce the number of required ports. Overall EOLE possesses a register file whose complexity is on-par with that of a regular wider-issue superscalar while the out-of-order components (scheduler, bypass) are greatly simplified. Moreover, thanks to Value Prediction, speedup is obtained on many benchmarks of the SPEC'00/'06 suite.

*6.2.1.3. Skewed Compressed Caches*

**Participant:** André Seznec.

Cache compression seeks the benefits of a larger cache with the area and power of a smaller cache. Ideally, a compressed cache increases effective capacity by tightly compacting compressed blocks, has low tag and metadata overheads, and allows fast lookups. Previous compressed cache designs, however, fail to achieve all these goals. In this study, we propose the Skewed Compressed Cache (SCC), a new hardware compressed cache that lowers overheads and increases performance. SCC tracks super- blocks to reduce tag overhead, compacts blocks into a variable number of sub-blocks to reduce internal fragmentation, but retains a direct tag-data mapping to find blocks quickly and eliminate extra metadata (i.e., no backward pointers). SCC does this using novel sparse super-block tags and a skewed associative mapping that takes compressed size into account. In our experiments, SCC provides on average 8% (up to 22%) higher performance, and on average 6% (up to 20%) lower total energy, achieving the benefits of the recent Decoupled Compressed Cache [47] with a factor of 4 lower area overhead and lower design complexity.

*This study was done in collaboration with Somayeh Sardashti and David Wood from University of Wisconsin.*

*6.2.1.4. Efficient Execution on Guarded Instruction Sets*

**Participant:** André Seznec.

ARM ISA based processors are no longer low complexity processors. Nowadays, ARM ISA based processor manufacturers are struggling to implement medium-end to high-end processor cores which implies implementing a state-of-the-art out-of-order execution engine. Unfortunately providing efficient out-of-order execution on legacy ARM codes may be quite challenging due to guarded instructions.

Predicting the guarded instructions addresses the main serialization impact associated with guarded instructions execution and the multiple definition problem. Moreover, guard prediction allows to use a global branch-and-guard history predictor to predict both branches and guards, often improving branch prediction accuracy. Unfortunately such a global branch-and-guard history predictor requires the systematic use of guard predictions. In that case, poor guard prediction accuracy would lead to poor overall performance on some applications.

Building on top of recent advances in branch prediction and confidence estimation, we propose a hybrid branch and guard predictor, combining a global branch history component and global branch-and-guard history component. The potential gain or loss due to the systematic use of guard prediction is dynamically evaluated at run-time. Two computing modes are enabled: systematic guard prediction use and high confidence only guard prediction use. Our experiments show that on most applications, an overwhelming majority of guarded instructions are predicted. Therefore a relatively inefficient but simple hardware solution can be used to execute the few unpredicted guarded instructions. Significant performance benefits are observed on most applications while applications with poorly predictable guards do not suffer from performance loss [7].

*This study is accepted to ACM Transactions on Architecture and Compiler Optimizations (to appear January 2015) and will be presented at the HIPEAC conference in January 2015.*

*6.2.1.5. Clustered microarchitecture*

**Participants:** Andrea Mondelli, Pierre Michaud, André Seznec.

In the last 10 years, the clock frequency of high-end superscalar processors did not increase significantly. Performance keeps being increased mainly by integrating more cores on the same chip and by introducing new instruction set extensions. However, this benefits only to some applications and requires rewriting and/or recompiling these applications. A more general way to increase performance is to increase the IPC, the number of instructions executed per cycle.

We argue that some of the benefits of technology scaling should be used to increase the IPC of future superscalar cores. Starting from microarchitecture parameters similar to recent commercial high-end cores, we show that an effective way to increase the IPC is to increase the issue width. But this must be done without impacting the clock cycle. We propose to combine two known techniques: clustering and register write specialization. The objective of past work on clustered microarchitecture was to allow a higher clock frequency while minimizing the IPC loss. This led researchers to consider narrow-issue clusters. Our objective, instead, is to increase the IPC without impacting the clock cycle, which means wide-issue clusters. We show that, on a wide-issue dual cluster, a very simple steering policy that sends 64 consecutive instructions to the same cluster, the next 64 instructions to the other cluster, and so on, permits tolerating an inter-cluster delay of several cycles. We also propose a method for decreasing the energy cost of sending results of one cluster to the other cluster.

This work is currently under submission.

*6.2.1.6. Adaptive Intelligent Memory Systems*

**Participants:** André Seznec, Aswinkumar Sridharan.

On multicores, the processors are sharing the memory hierarchy, buses, caches, and memory. The performance of any single application is impacted by its environment and the behavior of the other applications co-running on the multicore. Different strategies have been proposed to isolate the behavior of the different co-running applications, for example performance isolation cache partitioning, while several studies have addressed the global issue of optimizing throughput through the cache management.

However these studies are limited to a few cores (2-4-8) and generally feature mechanisms that cannot scale to 50-100 cores. Moreover so far the academic propositions have generally taken into account a single parameter, the cache replacement policy or the cache partitioning. Other parameters such as cache prefetching and its aggressiveness already impact the behavior of a single thread application on a uniprocessor. Cache prefetching policy of each thread will also impact the behavior of all the co-running threads.

Our objective is to define an Adaptive and Intelligent Memory System management hardware, AIMS. The goal of AIMS will be to dynamically adapt the different parameters of the memory hierarchy access for each individual co-running process in order to achieve a global objective such as optimized throughput, thread fairness or respecting quality of services for some privileged threads.

### *6.2.2. Microarchitecture Performance Modeling*

*6.2.2.1. Multiprogram throughput of multicore/SMT processors*

**Participant:** Pierre Michaud.

*This research was done in collaboration with Stijn Eyerman and Wouter Rogiest from Ghent University.*

There are several aspects to the performance of a multicore processor. One of them is multiprogram throughput, that is, how fast a multicore can execute several independent jobs. However, defining throughput metrics that are both meaningful and practical for computer architecture studies is not straightforward. We present a method to construct throughput metrics in a systematic way: we start by expressing assumptions on job sizes, job types distribution, scheduling, etc., that together define a theoretical throughput experiment. The throughput metric is then the average throughput of this experiment. Different assumptions lead to different metrics, so one should be aware of these assumptions when making conclusions based on results using a specific metric. Throughput metrics should always be defined from explicit assumptions, because this leads to a better understanding of the implications and limits of the results obtained with that metric. We elaborate multiple metrics based on different assumptions. In particular, we show that commonly used throughput metrics such as instructions per cycle and weighted speedup implicitly assume a variable workload, that is, a workload which depends on the machine being evaluated. However, in many situations, it is more realistic to assume a fixed workload. Hence we propose some new fixed-workload throughput metrics. Evaluating these new metrics requires to solve a continuous-time Markov chain. We released a software, TPCalc, that takes as input the performance results of individual coschedules simulations and computes fixed-workload throughput, taking advantage of multicore symmetries [15].

In a subsequent work, we applied our framework to symbiotic scheduling on a symmetric multicore or SMT processor. Symbiotic scheduling tries to exploit the fact, because of resource sharing (execution units, caches, memory bandwidth, etc.) and because different jobs have different characteristics, the performance may be increased by carefully choosing the coschedules. We show that, when assuming a fixed workload, an optimal schedule maximizing throughput can be found by solving a linear programming problem. However, the throughput gains we observed in our experiments, 3% on average, are significantly smaller than what we expected based on published studies on symbiotic scheduling. We analyzed the reasons for this and we found the two main reasons for this discrepancy: previous studies either did not consider a fixed workload but a variable one, or did not report throughput gains but response time reductions. Response time reductions can be artificially magnified by setting the job arrival rate close to the maximum throughput.

*This work will be presented at the ISPASS 2015 conference.*

*6.2.2.2. Modeling multi-threaded programs execution time in the many-core era*

**Participants:** Surya Khizakanchery Natarajan, Bharath Narasimha Swamy, André Seznec.

Estimating the potential performance of parallel applications on the yet-to-be-designed future many cores is very speculative. The traditional laws used to predict performance of an application do not reflect on the various scaling behaviour of a multi-threaded (MT) application leading to optimistic estimation of performance in manycore era. In this paper, we study the scaling behavior of MT applications as a function of input workload size and the number of cores. For some MT applications in the benchmark suites we analysed, our study shows that the serial fraction in the program increases with input workload size and can be a

scalability-limiting factor. Similar to previous studies [41] , we find that using a powerful core (heterogeneous architecture) to execute this serial part of the program can mitigate the impact of serial scaling and improve the overall performance of an application in many-core era [25].

### 6.2.3. Hardware/Software Approaches

*6.2.3.1. Helper threads*

**Participants:** Bharath Narasimha Swamy, Alain Ketterlin, André Seznec.

Heterogeneous Many Cores (HMC) architectures that mix many simple/small cores with a few complex/large cores are emerging as a design alternative that can provide both fast sequential performance for single threaded workloads and power-efficient execution for throughput oriented parallel workloads. The availability of many small cores in a HMC presents an opportunity to utilize them as low-power helper cores to accelerate memory-intensive sequential programs mapped to a large core. However, the latency overhead of accessing small cores in a loosely coupled system limits their utility as helper cores. Also, it is not clear if small cores can execute helper threads sufficiently in advance to benefit applications running on a larger, much powerful, core. In [24], we present a hardware/software framework called core-tethering to support efficient helper threading on heterogeneous many-cores. Core-tethering provides a co-processor like interface to the small cores that (a) enables a large core to directly initiate and control helper execution on the helper core and (b) allows efficient transfer of execution context between the cores, thereby reducing the performance overhead of accessing small cores for helper execution. Our evaluation on a set of memory intensive programs chosen from the standard benchmark suites show that, helper threads using moderately sized small cores can significantly accelerate a larger core compared to using a hardware prefetcher alone. We find that a small core provides a good trade-off against using an equivalent large core to run helper threads in a HMC. Additionally, helper prefetching on small cores when used along with hardware prefetching, can provide an alternate design point to growing instruction window size for achieving higher sequential performance on memory intensive applications.

*6.2.3.2. Branch Prediction and Performance of Interpreter*

**Participants:** Erven Rohou, André Seznec, Bharath Narasimha Swamy.

Interpreters have been used in many contexts. They provide portability and ease of development at the expense of performance. The literature of the past decade covers analysis of why interpreters are slow, and many software techniques to improve them. A large proportion of these works focuses on the dispatch loop, and in particular on the implementation of the switch statement: typically an indirect branch instruction. Folklore attributes a significant penalty to this branch, due to its high misprediction rate. We revisit this assumption, considering state-of-the-art branch predictors and the three most recent Intel processor generations on current interpreters. Using both hardware counters on Haswell, the latest Intel processor generation, and simulation of the ITTAGE, we show that the accuracy of indirect branch prediction is no longer critical for interpreters. We further compare the characteristics of these interpreters and analyze why the indirect branch is less important than before.

*This study [8] has been accepted for publication at CGO 2015 (International Symposium on Code Generation and Optimization).*

*6.2.3.3. Augmenting superscalar architecture for efficient many-thread parallel execution*

**Participants:** Sylvain Collange, André Seznec, Sajith Kalathingal.

We aim at exploring the design of a unique core that efficiently runs both sequential and massively parallel sections. We explore how the architecture of a complex superscalar core has to be modified or enhanced to efficiently run several threads from the same application.

Rather than vectorize at compile-time, our approach is to dynamically vectorize SPMD programs at the micro-architectural level. The SMT-SIMD hybrid core we propose extracts data parallelism from thread parallelism by scheduling groups of threads in lockstep, in a way inspired by the execution model of GPUs. As in GPUs, conditional branches whose outcomes differ between threads are handled with conditionally masked execution. However, while GPUs rely on explicit re-convergence instructions to restore lockstep execution, we target existing general-purpose instruction sets, in order to run legacy binary programs. Thus, the main challenge consists in detecting re-convergence points dynamically.

To handle this difficulty, we can build on [17]. In this work done in collaboration with Fernando Pereira and his team at UFMG, Brasil, we proposed instruction fetch policies that apply heuristics to maximize the cycles spent in lockstep execution, and evaluated them under a micro-architecture independent model [17]. Results highlight the necessity of a tradeoff between maximizing throughput and extracting data-level parallelism with lockstep execution.

## 6.3. Compiler, vectorization, interpretation

**Participants:** Erven Rohou, Emmanuel Riou, Bharath Narasimha Swamy, Arjun Suresh, André Seznec, Nabil Hallou, Alain Ketterlin, Sylvain Collange.

### *6.3.1. Compilers for emerging throughput architectures*

**Participant:** Sylvain Collange.

*This work is done in collaboration with Fernando Pereira and his team at UFMG, Brasil.*

GPU architectures present new challenges for compilers. Their performance characteristics demand SPMD programs with a high control-flow and memory regularity. Such architecture takes advantage of the regularity in programs to exploit data-level parallelism. In addition to the traditional challenges of code parallelization, new compilers for GPU and future throughput architectures face the task of improving the regularity of parallel programs. In particular, compiler analyses that identify control-flow divergence and memory divergence are a stepping stone for many optimizations. These optimizations include traditional code transformation such as loop interchange and tiling, which use divergence as an additional decision criterion, but also new optimizations specific to GPU architectures such as iteration delaying or branch fusion. In addition, the regularity parameter is an important aspect for workload characterization, as it provides a criterion for task scheduling in heterogeneous environments, such as multi-core processors with GPU. Our objectives include both accurate static and dynamic analyses for thread divergence, and the applications that it enables. We propose to combine static analyses with runtime checks, in order to get the best from both complementary approaches.

### *6.3.2. Improving sequential performance through memoization*

**Participants:** Erven Rohou, André Seznec, Arjun Suresh.

Many applications perform repetitive computations, even when properly programmed and optimized. Performance can be improved by caching results of pure functions, and retrieving them instead of recomputing a result (a technique called memoization).

We proposed a simple technique for enabling software memoization of any dynamically linked pure function and we illustrate our framework using a set of computationally expensive pure functions – the transcendental functions.

Our technique does not need the availability of source code and thus can be applied even 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 tool-chains, and can work with a variety of current architectures.

We present experimental results for x86-64 platform using both gcc and icc compiler tool-chains, and for ARM cortex-A9 platform using gcc. Our experiments include a mix of real world programs and standard benchmark suites: SPEC and Splash2x. On standard benchmark applications that extensively call the transcendental functions we report memoization benefits of upto 16 %, while much higher gains were realized for programs that call the expensive Bessel functions. Memoization was also able to regain a performance loss of 76 % in *bwaves* due to a known performance bug in the gcc libm implementation of *pow* function.

### *6.3.3. Code Obfuscation*

**Participant:** Erven Rohou.

*This research is done in collaboration with the group of Prof. Ahmed El-Mahdy at E-JUST, Alexandria, Egypt.*

A new obfuscation technique [27] based of decomposition of CFGs into threads has been proposed. We exploit the mainstream multi-core processing in these systems to substantially increase the complexity of programs, making reverse engineering more complicated. The novel method automatically partitions any serial thread into an arbitrary number of parallel threads, at the basic-block level. The method generates new control-flow graphs, preserving the blocks' serial successor relations and guaranteeing that one basic-block is active at a time through using guards. The method generates $m^n$ different combinations for $m$ threads and $n$ basic-blocks, significantly complicating the execution state. We also provide proof of correctness for the method.

We propose to leverage JIT compilation to make software tamper-proof. The idea is to constantly generate different versions of an application, even while it runs, to make reverse engineering hopeless. More precisely a JIT engine is used to generate new versions of a function each time it is invoked, applying different optimizations, heuristics and parameters to generate diverse binary code. A strong random number generator will guarantee that generated code is not reproducible, though the functionality is the same.

This work has been accepted for publication in January 2015 at the International Workshop on Dynamic Compilation Everywhere (DCE-2015).

### *6.3.4. Padrone*

**Participants:** Erven Rohou, Alain Ketterlin, Emmanuel Riou.

The objective of the ADT PADRONE is to design and develop a platform for re-optimization of binary executables at run-time. Development is ongoing, and an early prototype is functional. In [30], we described the infrastructure of Padrone, and showed that its profiling overhead is minimum. We illustrated its use through two examples. The first example shows how a user can easily write a tool to identify hotspots in their application, and how well they perform (for example, by computing the number of executed instructions per cycle). In the second example, we illustrate the replacement of a given function (typically a hotspot) by an optimized version, while the program runs.

We believe PADRONE fills an empty design point in the ecosystem of dynamic binary tools.

### *6.3.5. Dynamic Binary Re-vectorization*

**Participants:** Erven Rohou, Nabil Hallou, Alain Ketterlin, Emmanuel Riou.

*This work is done in collaboration with Philippe Clauss (Inria CAMUS).*

Applications are often under-optimized for the hardware on which they run. Several reasons contribute to this unsatisfying situation, including the use of legacy code, commercial code distributed in binary form, or deployment on compute farms. In fact, backward compatibility of instruction sets guarantees only the functionality, not the best exploitation of the hardware. In particular SIMD instruction sets are always evolving.

We proposed a runtime re-vectorization platform that dynamically adapts applications to execution hardware. Programs distributed in binary forms are re-vectorized at runtime for the underlying execution hardware. Focusing on the x86 SIMD extensions, we are able to automatically convert loops vectorized for SSE into the more recent and powerful AVX. A lightweight mechanism leverages the sophisticated technology put in a static vectorizer and adjusts, at minimal cost, the width of vectorized loops. We achieve speedups in line with a native compiler targeting AVX. Our re-vectorizer is implemented inside a dynamic optimization platform; its usage is completely transparent to the user and requires neither access to source code nor rewriting binaries.

## 6.4. WCET estimation

**Participants:** Damien Hardy, Hanbing Li, Isabelle Puaut, Erven Rohou.

Predicting the amount of resources required by embedded software is of prime importance for verifying that the system will fulfill its real-time and resource constraints. A particularly important point in hard real-time embedded systems is to predict the Worst-Case Execution Times (WCETs) of tasks, so that it can be proven that tasks temporal constraints (typically, deadlines) will be met. Our research concerns methods for obtaining automatically upper bounds of the execution times of applications on a given hardware. Our new results this year are on (i) multi-core architectures (ii) WCET estimation for faulty architectures (iii) traceability of flow information in compilers for WCET estimation.

### 6.4.1. WCET estimation and its interactions with compilation

#### 6.4.1.1. On the comparison of deterministic and probabilistic WCET estimation techniques

**Participants:** Damien Hardy, Isabelle Puaut.

*This is joint work with Jaume Abella, Eduardo Quinones and Francisco J. Cazorla from Barcelona Supercomputing Center*

Several timing analysis techniques have been proposed to obtain Worst-Case Execution Time (WCET) estimates of applications running on a particular hardware. They can be classified into two classes of approaches: deterministic timing analysis techniques (DTA), that produce a unique WCET estimate, and probabilistic timing analysis techniques (PTA) that produce multiple WCET estimates with associated probabilities. Both approaches have their static (SDTA, SPTA) and measurement-based (MBDTA, MBPTA) variants. The lack of comparison figures among those techniques makes complex the selection of the most appropriate one.

This work [19] makes a first attempt towards comparing comprehensively SDTA, SPTA and MBPTA qualitatively and quantitatively, under different cache configurations implementing LRU and random replacement. We identify strengths and limitations of each technique depending on the characteristics of the program under analysis and the hardware platform, thus providing users with guidance on which approach to choose depending on their target application and hardware platform.

### 6.4.2. WCET estimation for architectures with faulty caches

**Participants:** Damien Hardy, Isabelle Puaut.

Technology scaling, used to increase performance, has the negative consequence of providing less reliable silicon primitives, resulting in an increase of the probability of failure of circuits, in particular for SRAM cells. While space redundancy techniques exist to recover from failures and provide fault-free chips, they will not be affordable anymore in the future due to their growing cost. Consequently, other approaches like fine grain disabling and reconfiguration of hardware elements (e.g. individual functional units or cache blocks) will become economically necessary. This fine-grain disabling will lead to degraded performance compared to a fault-free execution.

A common implicit assumption in all static worst-case execution time (WCET) estimation methods is that the target processor is not subject to faults. Their result is not safe anymore when using fine grain disabling of hardware components, which degrades performance.

In [16] a method that statically calculates a probabilistic WCET bound in the presence of permanent faults in instruction caches is provided. The method, from a given program, cache configuration and probability of cell failure, derives a probabilistic WCET bound. An essential benefit of our approach is that its probabilistic nature stems only from the probability associated with the presence of faults. By construction, the worst-case execution path cannot be missed, since it is determined using static analysis, extended to cope with permanent faults. This allows our method to be used in safety-critical real-time systems. The method is computationally tractable, since it avoids the exhaustive enumeration of all possible fault locations. Experimental results show that the proposed method accurately estimates WCETs in the presence of permanent faults compared to a method that explores all possible locations for faults. On the one hand, the proposed method allows to quantify the impact of permanent faults on WCET estimates for chips with a known probability of cell failure for the whole chip lifetime. On the other hand, and most importantly, our work can also be used in architectural exploration frameworks to select the most appropriate fault management mechanisms, for current and future chip designs.

### 6.4.3. Traceability of flow information for WCET estimation

**Participants:** Hanbing Li, Isabelle Puaut, Erven Rohou.

This research is part of the ANR W-SEPT project.

Control-flow information is mandatory for WCET estimation, to guarantee that programs terminate (e.g. provision of bounds for the number of loop iterations) but also to obtain tight estimates (e.g. identification of infeasible or mutually exclusive paths). Such flow information is expressed though annotations, that may be calculated automatically by program/model analysis, or provided manually.

The objective of this work is to address the challenging issue of the mapping and transformation of the flow information from high level down to machine code. In our recent work [21], we have proposed a framework to systematically transform flow information from source code to machine code.

The framework defines a set of formulas to transform flow information for standard compiler optimizations. Transforming the flow information is done within the compiler, in parallel with transforming the code. There thus is no guessing what flow information have become, it is transformed along with the code. The framework is general enough to cover all linear flow constraints and all typical optimizations implemented in modern compilers. Our implementation in the LLVM compiler shows that we can improve the WCET of Malardalen benchmarks by 60% in average (up to 86%) by turning on optimizations. We also provide new insight on the impact of existing optimizations on the WCET.

### 6.4.4. Verified WCET estimation

**Participant:** Isabelle Puaut.

*This is joint work with Andre Oliveira Maroneze, David Pichardie and Sandrine Blazy from the Celtique group at Inria/IRISA Rennes.*

Current WCET estimation tools, even when based on sound static analysis techniques, are not verified. This may lead to bugs being accidentally introduced in the implementation. The main contribution of this work [13], [26] is a formally verified WCET estimation tool operating over C code.

Our tool is integrated to the formally verified CompCert C compiler. It is composed of two main parts: a loop bound estimation and an Implicit Path Enumeration Technique (IPET)-based WCET calculation method. We evaluated the precision of the WCET estimates on a reference benchmark and obtained results which are competitive with state-of-the-art WCET estimation techniques. The code of our tool is automatically generated from its formal specification. Furthermore, machine-checked proofs ensure the estimated WCET is at least as large as the actual WCET.

## 6.5. Computer arithmetic

**Participant:** Sylvain Collange.

### 6.5.1. Application-specific number systems

Collaboration with Mark G. Arnold, XLNS Research, USA.

Reconfigurable FPGA platforms let designers build efficient application-specific circuits, when the performance or energy efficiency of general-purpose CPUs is insufficient, and the production volume is not enough to offset the very high cost of building a dedicated integrated circuit (ASIC). One way to take advantage of the flexibility offered by FPGAs is to tailor arithmetic operators for the application. In particular, the Logarithmic Number System (LNS) is suitable for embedded applications dealing with low-precision, high-dynamic range numbers.

Like floating-point, LNS can represent numbers from a wide dynamic range with constant relative accuracy. However, while standard floating-point offer so-called subnormal numbers to represent numbers close to zero with constant absolute accuracy, LNS numbers abruptly overflow to zero, resulting in a gap in representable numbers close to zero that can impact the accuracy of numerical algorithms.

In collaboration with Mark G. Arnold, Sylvain Collange proposed a generalization of LNS that incorporates features analogous to subnormal floating-point [14]. The Denormal LNS (DLNS) system we introduce defines a class of hybrid number systems that offer quasi-constant absolute accuracy close to zero and quasi-constant relative accuracy on larger numbers. These systems can be configured to range from pure LNS (constant relative accuracy) to fixed-point (constant absolute accuracy across the whole range).

### *6.5.2. Deterministic floating-point primitives for high-performance computing*

Parallel algorithms such as reduction are ubiquitous in parallel programming, and especially high-performance computing. Although these algorithms rely on associativity, they are use on floating-point data, on which operations are not associative. As a result, computations become non-deterministic, and the result may change according to static and dynamic parameters such as machine configuration or task scheduling.

In collaboration with David Defour (UPVD), Stef Graillat and Roman Iakymchuk (LIP6), we introduce a solution to compute deterministic sums of floating-point numbers efficiently and with the best possible accuracy. A multi-level algorithm incorporating a filtering stage that uses fast vectorized floating-point expansions and an accumulation stage based on super-accumulators in a high-radix carry-save representation guarantees accuracy to the last bit even on degenerate cases while maintaining high performance in the common cases [35].

**ASAP Project-Team**

# 6. New Results

## 6.1. Highlights of the Year

- Anne-Marie Kermarrec is the recipient of the **ACM/IFIP/USENIX/Middleware 10-Years Best Paper Award**, for her paper *The peer sampling service: Experimental evaluation of unstructured gossip-based implementations* (Middleware 2004), co-authored with Márk Jelasity, Rachid Guerraoui, and Maarten van Steen.
- Anne-Marie Kermarrec is the recipient of the **WISE 2014 Best Paper Award**, for her paper [18], co-authored with Alexandra Olteanu and Karl Aberer.
- Michel Raynal is the recipient of the **PODC 2014 Best Paper Award**, for his paper [34], co-authored with Achour Mostefaoui and Moumen Hamouna.
- The MEDIEGO recommendation engine was demonstrated at **Le Web 14** in partnership with FranceTV.

BEST PAPERS AWARDS :
[18] **15th International Conference on Web Information System Engineering (WISE 2014)**. O. ALEXANDRA, A.-M. KERMARREC, K. ABERER.
[34] **ACM PODC**. A. MOSTEFAOUI, M. HAMOUNA, M. RAYNAL.

## 6.2. Models and abstractions for distributed systems

### *6.2.1. Signature-free asynchronous Byzantine consensus*

**Participant:** Michel Raynal.

In [34] we present a new round-based asynchronous consensus algorithm that copes with up to $t < n/3$ Byzantine processes, where $n$ is the total number of processes. In addition of not using signature, not assuming a computationally-limited adversary, while being optimal with respect to the value of $t$, this algorithm has several noteworthy properties: the expected number of rounds to decide is four, each round is composed of two or three communication steps and involves $O(n^2)$ messages, and a message is composed of a round number plus a single bit. To attain this goal, the consensus algorithm relies on a common coin as defined by Rabin, and a new extremely simple and powerful broadcast abstraction suited to binary values. The main target when designing this algorithm was to obtain a cheap and simple algorithm. This was motivated by the fact that, among the first-class properties, simplicity –albeit sometimes under-estimated or even ignored– is a major one.

This is a joint work with Achour Mostéfaouin and Hamouma Moumen. It received the PODC 2014 Best Paper Award.

### *6.2.2. Randomized mutual exclusion with constant amortized RMR complexity on the DSM*

**Participant:** George Giakkoupis.

In [30] we settle an open question by determining the remote memory reference (RMR) complexity of randomized mutual exclusion, on the distributed shared memory model (DSM) with atomic registers, in a weak but natural (and stronger than oblivious) adversary model. In particular, we present a mutual exclusion algorithm that has constant expected amortized RMR complexity and is deterministically deadlock free. Prior to this work, no randomized algorithm with $o(\log n/\log\log n)$ RMR complexity was known for the DSM model. Our algorithm is fairly simple, and compares favorably with one by Bender and Gilbert (FOCS 2011) for the CC model, which has expected amortized RMR complexity $O(\log^2 \log n)$ and provides only probabilistic deadlock freedom.

This is a joint work with Philipp Woelfel (Univ. of Calgary, Canada).

### 6.2.3. Reliable shared memory abstraction on top of asynchronous Byzantine message-passing systems

**Participants:** Michel Raynal, Julien Stainer.

This work is on the construction and the use of a shared memory abstraction on top of an asynchronous message-passing system in which up to t processes may commit Byzantine failures. This abstraction consists of arrays of n single-writer/multi-reader atomic registers, where $n$ is the number of processes. Differently from usual atomic registers which record a single value, each of these atomic registers records the whole history of values written to it. A distributed algorithm building such a shared memory abstraction it first presented. This algorithm assumes $t < n/3$, which is shown to be a necessary and sufficient condition for such a construction. Hence, the algorithm is resilient-optimal. Then we present distributed algorithms built on top of this shared memory abstraction, which cope with up to $t$ Byzantine processes. The simplicity of these algorithms constitutes a strong motivation for such a shared memory abstraction in the presence of Byzantine processes. For a lot of problems, algorithms are more difficult to design and prove correct in a message-passing system than in a shared memory system. Using a protocol stacking methodology, the aim of the proposed abstraction is to allow an easier design (and proof) of distributed algorithms, when the underlying system is an asynchronous message-passing system prone to Byzantine failures.

This work was done in collaboration with Damien Imbs and Sergio Rajsbaum. It has been published in SIRROCCO [32] and as a technical report [43].

### 6.2.4. Distributed Universality

**Participants:** Michel Raynal, Julien Stainer.

A notion of a universal construction suited to distributed computing has been introduced by M. Herlihy in his celebrated paper "Wait-free synchronization" (ACM TOPLAS, 1991). A universal construction is an algorithm that can be used to wait-free implement any object defined by a sequential specification. Herlihy's paper shows that the basic system model, which supports only atomic read/write registers, has to be enriched with consensus objects to allow the design of universal constructions. The generalized notion of a $k$-universal construction has been recently introduced by Gafni and Guerraoui (CONCUR 2011). A $k$-universal construction is an algorithm that can be used to simultaneously implement $k$ objects (instead of just one object), with the guarantee that at least one of the $k$ constructed objects progresses forever. While Herlihy's universal construction relies on atomic registers and consensus objects, a $k$-universal construction relies on atomic registers and $k$-simultaneous consensus objects (which are wait-free equivalent to $k$-set agreement objects in the read/write system model). This work significantly extends the universality results introduced by Herlihy and Gafni-Guerraoui. In particular, we present a $k$-universal construction which satisfies the following five desired properties, which are not satisfied by the previous $k$-universal construction: (1) among the $k$ objects that are constructed, at least $\ell$ objects (and not just one) are guaranteed to progress forever; (2) the progress condition for processes is wait-freedom, which means that each correct process executes an infinite number of operations on each object that progresses forever; (3) if any of the $k$ constructed objects stops progressing, all its copies (one at each process) stop in the same state; (4) the proposed construction is contention-aware, in the sense that it uses only read/write registers in the absence of contention; and (5) it is generous with respect to the obstruction-freedom progress condition, which means that each process is able to complete any one of its pending operations on the $k$ objects if all the other processes hold still long enough. The proposed construction, which is based on new design principles, is called a ($k$,$\ell$)-universal construction. It uses a natural extension of $k$-simultaneous consensus objects, called ($k$,$\ell$)-simultaneous consensus objects (($k$,$\ell$)-SC). Together with atomic registers, ($k$,$\ell$)-SC objects are shown to be necessary and sufficient for building a ($k$,$\ell$)-universal construction, and, in that sense, ($k$,$\ell$)-SC objects are ($k$,$\ell$)-universal.

This work was done in collaboration with Gadi Taubenfeld. It has been published as a brief announcement in PODC [37] and the full version appeared in OPODIS [38]. A version has also been published as a technical report [45].

### 6.2.5. Computing in the presence of concurrent solo executions

**Participants:** Michel Raynal, Julien Stainer.

In a wait-free model any number of processes may crash. A process runs solo when it computes its local output without receiving any information from other processes, either because they crashed or they are too slow. While in wait-free shared-memory models at most one process may run solo in an execution, any number of processes may have to run solo in an asynchronous wait-free message-passing model. This work is on the computability power of models in which several processes may concurrently run solo. It first introduces a family of round-based wait-free models, called the d-solo models, $1 \leq d \leq n$, where up to $d$ processes may run solo. We then give a characterization of the colorless tasks that can be solved in each $d$-solo model. We also introduce the ($d$,$\epsilon$)-solo approximate agreement task, which generalizes $\epsilon$-approximate agreement, and proves that ($d$,$\epsilon$)-solo approximate agreement can be solved in the $d$-solo model, but cannot be solved in the $(d+1)$-solo model. We study also the relation linking $d$-set agreement and ($d$,$\epsilon$)-solo approximate agreement in asynchronous wait-free message-passing systems. These results establish for the first time a hierarchy of wait-free models that, while weaker than the basic read/write model, are nevertheless strong enough to solve non-trivial tasks.

This work was done in collaboration with Maurice Herlihy and Sergio Rajsbaum. It has been published in LATIN [31].

### 6.2.6. A simple broadcast algorithm for recurrent dynamic systems

**Participants:** Michel Raynal, Julien Stainer.

This work presents a simple broadcast algorithm suited to dynamic systems where links can repeatedly appear and disappear. The algorithm is proved correct and a simple improvement is introduced, that reduces the number and the size of control messages. As it extends in a simple way a classical network traversal algorithm to the dynamic context, the proposed algorithm has also pedagogical flavor.

This work was done in collaboration with Jiannong Cao and Weigang Wu. It has been published in AINA [36].

### 6.2.7. Fisheye consistency: Keeping data in synch in a georeplicated world

**Participants:** Michel Raynal, François Taïani.

Over the last thirty years, numerous consistency conditions for replicated data have been proposed and implemented. Popular examples of such conditions include linearizability (or atomicity), sequential consistency, causal consistency, and eventual consistency. These consistency conditions are usually defined independently from the computing entities (nodes) that manipulate the replicated data; i.e., they do not take into account how computing entities might be linked to one another, or geographically distributed. To address this lack, as a first contribution, this work [41] introduces the notion of proximity graph between computing nodes. If two nodes are connected in this graph, their operations must satisfy a strong consistency condition, while the operations invoked by other nodes are allowed to satisfy a weaker condition. The second contribution is the use of such a graph to provide a generic approach to the hybridization of data consistency conditions into the same system. We illustrate this approach on sequential consistency and causal consistency, and present a model in which all data operations are causally consistent, while operations by neighboring processes in the proximity graph are sequentially consistent. The third contribution of this work is the design and the proof of a distributed algorithm based on this proximity graph, which combines sequential consistency and causal consistency (the resulting condition is called fisheye consistency). In doing so this work not only extends the domain of consistency conditions, but provides a generic provably correct solution of direct relevance to modern georeplicated systems.

This work was done in collaboration with Roy Friedman (The Technion, Haifa, Israel)

## 6.3. Large-scale and user-centric distributed systems

### 6.3.1. Archiving cold data in warehouses with clustered network coding

**Participants:** Fabien André, Anne-Marie Kermarrec.

Modern storage systems now typically combine plain replication and erasure codes to reliably store large amount of data in datacenters. Plain replication allows a fast access to popular data, while erasure codes, e.g., Reed-Solomon codes, provide a storage-efficient alternative for archiving less popular data. Although erasure codes are now increasingly employed in real systems, they experience high overhead during maintenance, i.e., upon failures, typically requiring files to be decoded before being encoded again to repair the encoded blocks stored at the faulty node.

In this work, we proposed a novel erasure code system, tailored for networked archival systems. The efficiency of our approach relies on the joint use of random codes and a clustered placement strategy. Our repair protocol leverages network coding techniques to reduce by 50% the amount of data transferred during maintenance, by repairing several cluster files simultaneously. We demonstrated both through an analysis and extensive experimental study conducted on a public testbed that our approach significantly decreases both the bandwidth overhead during the maintenance process and the time to repair lost data. We also showed that using a non-systematic code does not impact the throughput, and comes only at the price of a higher CPU usage. Based on these results, we evaluated the impact of this higher CPU consumption on different configurations of data coldness by determining whether the cluster's network bandwidth dedicated to repair or CPU dedicated to decoding saturates first.

This work has been conducted in collaboration with Erwan Le Merrer, Nicolas Le Scouarnec, Gilles Straub (Technicolor) and A. van Kempen (Univ. Nantes) and published in ACM Eurosys 2014 [19].

### 6.3.2. WebGC: Browser-based gossiping

**Participants:** Raziel Carvajal Gomez, Davide Frey, Anne-Marie Kermarrec.

The advent of browser-to-browser communication technologies like WebRTC has renewed interest in the peer-to-peer communication model. However, the available WebRTC code base still lacks important components at the basis of several peer-to-peer solutions. Through a collaboration with Mathieu Simonin from the Inria SED in the context of the Brow2Brow ADT project, we started to tackle this problem by proposing WebGC, a library for gossip-based communication between web browsers. Due to their inherent scalability, gossip-based, or epidemic protocols constitute a key component of a large number of decentralized applications. WebGC thus represents an important step towards their wider spread. We demonstrated a preliminary version of the library at Middleware 2014 [47].

### 6.3.3. Large-scale graph processing in datacenters with bandwidth guarantees

**Participants:** Nitin Chiluka, Anne-Marie Kermarrec.

Recent research has shown that the performance of data-intensive applications in multi-tenant datacenters can be severely impacted by each other's network usage. Starvation for network bandwidth in such datacenters typically results in significantly longer completion times for large-scale distributed applications. To address this concern, researchers propose bandwidth guarantees for all the virtual machines (VMs) initiated by each tenant in the datacenter in order to provide a predictable performance for their applications. In our work, we focus on large-scale graph processing in such datacenters. More specifically, given $k$ VMs with their respective bandwidth constraints and a large graph, we perform a $k$-way partition on the graph such that the subsequent computation of various algorithms (e.g., PageRank, graph factorization) take minimal time.

### 6.3.4. Scaling KNN computation over large graphs on a PC

**Participants:** Nitin Chiluka, Anne-Marie Kermarrec, Javier Olivares.

Frameworks such as GraphChi and X-Stream are increasingly gaining attention for their ability to perform scalable computation on large graphs by leveraging disk and memory on a single commodity PC. These frameworks rely on the graph structure to remain the same for the entire pe- riod of computation of various algorithms such as PageRank and triangle counting. As a consequence, these frameworks are not applicable to algorithms that require the graph struc- ture to change during their computation. In this work, we focus on one such algorithm – K-Nearest Neighbors (KNN) – which is widely used in recommender systems. Our approach aims to minimize random accesses to disk as well as the amount of data loaded/unloaded from/to disk so as to better utilize the computational power, thus improving the algorithmic efficiency. The preliminary design and results of our approach appeared in Middleware 2014 [23].

### 6.3.5. Privacy-preserving distributed collaborative filtering

**Participants:** Davide Frey, Arnaud Jégou, Anne-Marie Kermarrec.

In collaboration with Antoine Boutet from the Univ. St Etienne, and Rachid Guerraoui from EPFL, we proposed a new mechanism to preserve privacy while leveraging user profiles in distributed recommender systems. Our mechanism relies on two contributions: (*i*) an original obfuscation scheme, and (*ii*) a randomized dissemination protocol. We showed that our obfuscation scheme hides the exact profiles of users without significantly decreasing their utility for recommendation. In addition, we precisely characterized the conditions that make our randomized dissemination protocol differentially private.

We compared our mechanism with a non-private as well as with a fully private alternative. We considered a real dataset from a user survey and report on simulations as well as planetlab experiments. In short, our extensive evaluation showed that our twofold mechanism provides a good trade-off between privacy and accuracy, with little overhead and high resilience.

### 6.3.6. Behave: Behavioral cache for web content

**Participants:** Davide Frey, Anne-Marie Kermarrec.

In collaboration with Mathieu Goessens, a former intern of the team, we proposed Behave: a novel approach for peer-to-peer cache-oriented applications such as CDNs. Behave relies on the principle of Behavioral Locality inspired from collaborative filtering. Users that have visited similar websites in the past will have local caches that provide interesting content for one another.

Behave exploits epidemic protocols to build overlapping communities of peers with similar interests. Peers in the same one-hop community federate their cache indexes in a Behavioral cache. Extensive simulations on a real data trace show that Behave can provide zero-hop lookup latency for about 50% of the content available in a DHT-based CDN. The results of this work were published at DAIS 2014 [26].

### 6.3.7. HyRec: Leveraging browsers for scalable recommenders

**Participants:** Davide Frey, Anne-Marie Kermarrec.

The ever-growing amount of data available on the Internet calls for personalization. Yet, the most effective personalization schemes, such as those based on collaborative filtering (CF), are notoriously resource greedy. In this work, we proposed HyRec, an online cost-effective scalable system for user-based CF personalization. HyRec offloads recommendation tasks onto the web browsers of users, while a server orchestrates the process and manages the relationships between user profiles.

We fully implemented HyRec and we extensively evaluated it on several workloads from MovieLens and Digg. Our experiments conveyed the ability of HyRec to reduce the operation costs of content providers by nearly 50% and to provide a 100-fold improvement in scalability with respect to a centralized (or cloud-based recommender approach), while preserving the quality of personalization. HyRec is also virtually transparent to users and induces only 3% of the bandwidth consumption of a p2p solution. This work was done in collaboration with Antoine Boutet from the Univ. St Etienne, as well as with Rachid Guerraoui, and Rhicheek Patra from EPFL. It resulted in a publication at Middleware 2014 [22].

### 6.3.8. Landmark-based similarity for p2p collaborative filtering

**Participants:** Davide Frey, Anne-Marie Kermarrec, Antoine Rault, François Taïani.

Computing $k$-nearest-neighbor graphs constitutes a fundamental operation in a variety of data-mining applications. As a prominent example, user-based collaborative-filtering provides recommendations by identifying the items appreciated by the closest neighbors of a target user. As this kind of applications evolve, they will require KNN algorithms to operate on more and more sensitive data. This has prompted researchers to propose decentralized peer-to-peer KNN solutions that avoid concentrating all information in the hands of one central organization. Unfortunately, such decentralized solutions remain vulnerable to malicious peers that attempt to collect and exploit information on participating users.

We seek to overcome this limitation by proposing *H&S* (Hide & Share), a novel landmark-based similarity mechanism for decentralized KNN computation. Landmarks allow users (and the associated peers) to estimate how close they lay to one another without disclosing their individual profiles.

We evaluate *H&S* in the context of a user-based collaborative-filtering recommender with publicly available traces from existing recommendation systems. We show that although landmark-based similarity does disturb similarity values (to ensure privacy), the quality of the recommendations is not as significantly hampered. We also show that the mere fact of disturbing similarity values turns out to be an asset because it prevents a malicious user from performing a profile reconstruction attack against other users, thus reinforcing users' privacy. Finally, we provide a formal privacy guarantee by computing the expected amount of information revealed by *H&S* about a user's profile.

This work was done in collaboration with Jingjing Wang, and Rachid Guerraoui.

### 6.3.9. Adaptation for the masses: Towards decentralized adaptation in large-scale p2p recommenders

**Participants:** Davide Frey, Anne-Marie Kermarrec, François Taïani.

Decentralized recommenders have been proposed to deliver privacy-preserving, personalized and highly scalable on-line recommendation services. Current implementations tend, however, to rely on hard-wired, mechanisms that cannot adapt. Deciding beforehand which hard-wired mechanism to use can be difficult, as the optimal choice might depend on conditions that are unknown at design time. In [27], we have proposed a framework to develop dynamically adaptive decentralized recommendation systems. Our proposal supports a decentralized form of adaptation, in which individual nodes can independently select, and update their own recommendation algorithm, while still collectively contributing to the overall system's services.

This work was done in collaboration with Christopher Maddock and Andreas Mauthe (Univ. of Lancaster, UK).

### 6.3.10. Tight bounds for rumor spreading with vertex expansion

**Participant:** George Giakkoupis.

In [28] we establish an upper bound for the classic PUSH-PULL rumor spreading protocol on general graphs, in terms of the vertex expansion of the graph. We show that $O(\log^2(n)/\alpha)$ rounds suffice with high probability to spread a rumor from any single node to all n nodes, in any graph with vertex expansion at least $\alpha$. This bound matches a known lower bound, and settles the natural question on the relationship between rumor spreading and vertex expansion asked by Chierichetti, Lattanzi, and Panconesi (SODA 2010). Further, some of the arguments used in the proof may be of independent interest, as they give new insights, for example, on how to choose a small set of nodes in which to plant the rumor initially, to guarantee fast rumor spreading.

### 6.3.11. Greedy routing in small-world networks with power-law degrees

**Participant:** George Giakkoupis.

In [12] we study decentralized routing in small-world networks that combine a wide variation in node degrees with a notion of spatial embedding. Specifically, we consider a variant of J. Kleinberg's grid-based small-world model in which (1) the number of long-range edges of each node is not fixed, but is drawn from a power-law probability distribution with exponent parameter $\alpha \geq 0$ and constant mean, and (2) the long-range edges are considered to be bidirectional for the purposes of routing. This model is motivated by empirical observations indicating that several real networks have degrees that follow a power-law distribution. The measured power-law exponent $\alpha$ for these networks is often in the range between 2 and 3. For the small-world model we consider, we show that when $2 < \alpha < 3$ the standard greedy routing algorithm, in which a node forwards the message to its neighbor that is closest to the target in the grid, finishes in an expected number of $O(\log^{\alpha-1} n \cdot \log\log n)$ steps, for any source–target pair. This is asymptotically smaller than the $O(\log^2 n)$ steps needed in Kleinberg's original model with the same average degree, and approaches $O(\log n)$ as $\alpha$ approaches 2. Further, we show that when $0 \leq \alpha < 2$ or $\alpha \geq 3$ the expected number of steps is $O(\log^2 n)$, while for $\alpha = 2$ it is $O(\log^{4/3} n)$. We complement these results with lower bounds that match the upper bounds within at most a $\log\log n$ factor.

This is a joint work with Pierre Fraigniaud (Inria Paris-Rocquencourt and CNRS).

### 6.3.12. Randomized rumor spreading in dynamic graphs

**Participant:** George Giakkoupis.

In [29] we consider the well-studied rumor spreading model in which nodes contact a random neighbor in each round in order to push or pull the rumor. Unlike most previous works which focus on static topologies, we look at a dynamic graph model where an adversary is allowed to rewire the connections between vertices before each round, giving rise to a sequence of graphs, $G_1, G_2, \ldots$ Our first result is a bound on the rumor spreading time in terms of the conductance of those graphs. We show that if the degree of each node does not change much during the protocol (that is, by at most a constant factor), then the spread completes within $t$ rounds for some $t$ such that the sum of conductances of the graphs $G_1$ up to $G_t$ is $O(\log n)$. This result holds even against an adaptive adversary whose decisions in a round may depend on the set of informed vertices before the round, and implies the known tight bound with conductance for static graphs. Next we show that for the alternative expansion measure of vertex expansion, the situation is different. An adaptive adversary can delay the spread of rumor significantly even if graphs are regular and have high expansion, unlike in the static graph case where high expansion is known to guarantee fast rumor spreading. However, if the adversary is oblivious, i.e., the graph sequence is decided before the protocol begins, then we show that a bound close to the one for the static case holds for any sequence of regular graphs.

This is a joint work with Thomas Sauerwald (Univ. of Cambridge, UK) and Alexandre Stauffer (Univ. of Bath, UK).

### 6.3.13. Privacy-preserving dissemination in social networks and microblogs

**Participants:** George Giakkoupis, Arnaud Jégou, Anne-Marie Kermarrec, Nupur Mittal.

Online micro-blogging services and social networks, as exemplified by Twitter and Facebook, have emerged as an important means of disseminating information quickly and at large scale. A standard mechanism in micro-blogging that allows for interesting content to reach a wider audience is that of *reposting* (i.e., *retweeting* in Twitter, or *sharing* in Facebook) of content initially posted by another user. Motivated by recent events in which users were prosecuted merely for reposting anti-government information, we present in [42] Riposte, a randomized reposting scheme that provides privacy guarantees against such charges. The idea is that if the user likes a post, Riposte will repost it only with some (carefully chosen) probability; and if the user does not like it, Riposte may still repost it with a slightly smaller probability. These probabilities are computed for each user as a function of the number of connections of the user in the network, and the extent to which the post has already reached those connections. The choice of these probabilities is based on results for branching processes, and ensures that interesting posts (liked by a large fraction of users) are likely to disseminate widely, whereas uninteresting posts (or spam) do not spread. Riposte is executed locally at the user, thus the user's opinion on the post is not communicated to the micro-blogging server. In this work, we quantify Riposte's ability to protect users in terms of differential privacy and provide analytical bounds on the dissemination of posts. We also do extensive experiments based on topologies of real networks, including Twitter, Facebook, Renren, Google+ and LiveJournal.

This work has been carried out in collaboration with Rachid Guerraoui (EPFL).

### 6.3.14. Adaptive streaming

**Participants:** Ali Gouta, Anne-Marie Kermarrec.

HTTP Adaptive Streaming (HAS) is gradually being adopted by Over The Top (OTT) content providers. In HAS, a wide range of video bitrates of the same video content are made available over the internet so that clients' players pick the video bitrate that best fit their bandwidth. Yet, this affects the performance of some major components of the video delivery chain, namely CDNs or transparent caches since several versions of the same content compete to be cached. In this context, we investigated the benefits of a Cache Friendly HAS system (CF-DASH), which aims to improve the caching efficiency in mobile networks and to sustain the quality of experience of mobile clients. We conducted our work by presenting a set of observations we made

on a large number of clients requesting HAS contents. We introduced the CF-Dash system and our testbed implementation. Finally, we evaluated CF-dash based on trace-driven simulations and testbed experiments. Our validation results are promising. Simulations on real HAS traffic show that we achieve a significant gain in hit-ratio that ranges from 15% up to 50%. This work was done in collaboration with Zied Aouini, Yannick Le Louedec and Diallo Mamadou, and was published in NOSSDAV 2014 [39].

### 6.3.15. Predictive capabilities of social and interest affinity for recommendations

**Participant:** Anne-Marie Kermarrec.

The advent of online social networks created new prediction opportunities for recommender systems: instead of relying on past rating history through the use of collaborative filtering (CF), they can leverage the social relations among users as a predictor of user tastes similarity. Alas, little effort has been put into understanding when and why (e.g., for which users and what items) the social affinity (i.e., how well connected users are in the social network) is a better predictor of user preferences than the interest affinity among them as algorithmically determined by CF, and how to better evaluate recommendations depending on, for instance, what type of users a recommendation application targets. This overlook is explained in part by the lack of a systematic collection of datasets including both the explicit social network among users and the collaborative annotated items. In this work, we conducted an extensive empirical analysis on six real-world publicly available datasets, which dissects the impact of user and item attributes, such as the density of social ties or item rating patterns, on the performance of recommendation strategies relying on either the social ties or past rating similarity. Our findings represent practical guidelines that can assist in future deployments and mixing schemes. This work has been done in collaboration with Karl Aberer and Alexandra Olteanu (EPFL Swizerland). The paper received the Best Paper Award at the WISE International Conference [18].

### 6.3.16. Polystyrene: The decentralized data shape that never dies

**Participants:** Anne-Marie Kermarrec, François Taïani.

Decentralized topology construction protocols organize nodes along a predefined topology (e.g. a torus, ring, or hypercube). Such topologies have been used in many contexts ranging from routing and storage systems, to publish-subscribe and event dissemination. Since most topologies assume no correlation between the physical location of nodes and their positions in the topology, they do not handle catastrophic failures well, in which a whole region of the topology disappears. When this occurs, the overall shape of the system typically gets lost. This is highly problematic in applications in which overlay nodes are used to map a virtual data space, be it for routing, indexing or storage. In this work [20], we propose a novel decentralized approach that maintains the initial shape of the topology even if a large (consecutive) portion of the topology fails. Our approach relies on the dynamic decoupling between physical nodes and virtual ones enabling a fast reshaping. For instance, our results show that a 51,200-node torus converges back to a full torus in only 10 rounds after 50% of the nodes have crashed. Our protocol is both simple and flexible and provides a novel form of collective survivability that goes beyond the current state of the art.

This work has been done in collaboration with Simon Bouget (ENS Rennes) and Hoel Kervadec (INSA Rennes).

### 6.3.17. Link-prediction for very large scale graphs using distributed graph engines

**Participants:** Anne-Marie Kermarrec, François Taïani, Juan Manuel Tirado Martin.

In this project, we consider how the emblematic problem of link-prediction can be implemented efficiently in gather-apply-scatter (GAS) platforms, a popular distributed graph-computation model. Our proposal, called SNAPLE, exploits a novel highly-localized vertex scoring technique, and minimizes the cost of data flow while maintaining prediction quality. When used within GraphLab, SNAPLE can scale to extremely large graphs that a standard implementation of link prediction on cannot handle within the same platform. More precisely, we show that our approach can process a graph containing 1.4 billions edges on a 256 cores cluster in less than three minutes, with no penalty in the quality of predictions. This result corresponds to an over-linear speedup of 30 against a 20-core stand-alone machine running a non-distributed state-of-the-art solution.

### 6.3.18. GOSSIPKIT: A unified component framework for gossip

**Participant:** François Taïani.

Although the principles of gossip protocols are relatively easy to grasp, their variety can make their design and evaluation highly time consuming. This problem is compounded by the lack of a unified programming framework for gossip, which means developers cannot easily reuse, compose, or adapt existing solutions to fit their needs, and have limited opportunities to share knowledge and ideas. In [17], we have considered how component frameworks, which have been widely applied to implement middleware solutions, can facilitate the development of gossip-based systems in a way that is both generic and simple. We show how such an approach can maximise code reuse, simplify the implementation of gossip protocols, and facilitate dynamic evolution and re-deployment.

This work was done in collaboration with Shen Lin (SAP Labs) and Gordon Blair (Univ. of Lancaster, UK).

### 6.3.19. Towards a new model for cyber foraging

**Participant:** François Taïani.

Cyber foraging seeks to expand the capabilities and battery life of mobile devices by offloading intensive computations to nearby computing nodes (the surrogates). Although promising, current approaches to cyber foraging tend to impose a strict separation between the application state maintained on the mobile device, and data processed on the surrogates. In [33], we argue that this separation limits the applicability of cyber foraging, and explore how state sharing could be implemented in practice.

This work was done in collaboration with Diogo Lima and Hugo Miranda (Univ. of Lisbon, Portugal).

**ASCOLA Project-Team**

# 6. New Results

## 6.1. Highlights of the Year

Nicolas Tabareau was awarded a starting grant from the European Research Council (ERC), the most prestigious type of research projects of the European Union for young researchers. From 2015–2020 he will pursue research on "CoqHoTT: Coq for Homotopy Type Theory."

Jonathan Pastor has won the joint 1st prize at the Grid5000 Scale challenge, an international challenge for large-scale experiments on geographically-distributed cluster environments. Jonathan has shown with a colleague how to deploy and manage thousands of VMs in such an environment using his approach to fully distributed virtual machine management.

This year we have provided major research results in two domains. First, we have developed several new approaches for the formal reasoning over software in the domains of theorem proving [31], as well as reasoning over distributed interaction protocols [32] and software compositions [24]. Second, we have developed new methods supporting dynamic computations over the cloud, both by means of more elastic cloud applications [27] and better locality management for the dynamic placement of virtual machines in Cloud infrastructures [29].

## 6.2. Programming Languages

**Participants:** Ronan-Alexandre Cherrueau, Rémi Douence, Hervé Grall, Thomas Ledoux, Florent Marchand de Kerchove de Denterghem, Jacques Noyé, Jean-Claude Royer, Mario Südholt.

### *6.2.1. Formal Methods, logics and type theory*

This year we have published new results extending previous type theories: we have introduced a notion of universe polymorphism for the theorem prover Coq and new type-based mechanisms for the definition and analysis of program equivalences. We have also shown how to harness capabilities, well-known in the security domain, in the context of the functional programming language Haskell. These results are detailed in the current section.

Furthermore, we have applied formal methods and typing in the context of aspect oriented programming ([12], [16], [24]) and in the context of distributed programming (aspectual session types [32]). We have also developed a framework for the formal definition and analysis of accountability properties based on temporal logics. These different results are detailed in Sec. 6.3 for details.

#### *6.2.1.1. Universe Polymorphism in Coq*

Universes are used in type theory to ensure consistency by checking that definitions are well-stratified according to a certain hierarchy. In the case of the Coq proof assistant, based on the predicative Calculus of Inductive Constructions (pCIC), this hierarchy is built from an impredicative sort Prop and an infinite number of predicative Type universes. A cumulativity relation represents the inclusion order of universes in the core theory. Originally, universes were thought to be floating levels, and definitions to implicitly constrain these levels in a consistent manner. This works well for most theories, however the globality of levels and constraints precludes generic constructions on universes that could work at different levels. We have introduced universe polymorphism [31] that extends this setup by adding local bindings of universes and constraints, supporting generic definitions over universes, reusable at different levels. This provides the same kind of code reuse facilities as ML-style parametric polymorphism. However, the structure and hierarchy of universes is more complex than bare polymorphic type variables.

*6.2.1.2. A Logical Study of Program Equivalence*

Proving program equivalence for a functional language with references is a notoriously difficult problem. The goal of the thesis of Guilhem Jaber on "A Logical Study of Program Equivalence" [G. Jaber, Mines Nantes, July 14] was to propose a logical system in which such proofs can be formalized, and in some cases inferred automatically. In the first part, a generic extension method of dependent type theory has been proposed, based on a forcing interpretation seen as a presheaf translation of type theory. This extension equips type theory with guarded recursive constructions, which are subsequently used to reason on higher-order references. In the second part, he has defined a nominal game semantics for a language with higher-order references. It marries the categorical structure of game semantics with a trace representation of denotations of programs, which can be computed operationally and thus have good modularity properties. Using this semantics, he has proven completeness of Kripke logical relations defined in a direct way, using guarded recursive types, without using biorthogonality. The problem of contextual equivalence is then reduced to the satisfiability of an automatically generated formula defined in this logic, that is, to the existence of a world validating this formula. Under some conditions, this satisfiability can be decided using a SMT solver.

*6.2.1.3. Effect Capabilities For Haskell*

Computational effects complicate the tasks of reasoning about and maintaining software, due to the many kinds of interferences that can occur. While different proposals have been formulated to alleviate the fragility and burden of dealing with specific effects, such as state or exceptions, there is no prevalent robust mechanism that addresses the general interference issue. Building upon the idea of capability-based security, we have proposed effect capabilities [25] as an effective and flexible manner to control monadic effects and their interferences. Capabilities can be selectively shared between modules to establish secure effect-centric coordination. We have further refined capabilities with type-based permission lattices to allow fine-grained decomposition of authority. An implementation of effect capabilities in Haskell has been done, using type classes to establish a way to statically share capabilities between modules, as well as to check proper access permissions to effects at compile time.

### *6.2.2. Language Mechanisms*

In 2014, we have proposed new general language-based mechanisms for concurrent event-based systems and sequential programming languages. Moreover, we have investigated domain-specific languages that support aspect-oriented programming and provide control over propagation strategies in constraint solvers. These results are detailed in the remainder of this section.

Furthermore, we have proposed language support for the definition and enforcement of security properties, in particular related to the accountability of service-based systems, see Sec. 6.3 .

*6.2.2.1. Concurrent Event-Based Programming*

Advanced concurrency abstractions overcome the drawbacks of low-level techniques such as locks and monitors, freeing programmers that implement concurrent applications from the burden of concentrating on low-level details. However, with current approaches the coordination logic involved in complex coordination schemas is fragmented into several pieces including join patterns, data emissions triggered in different places of the application, and the application logic that implicitly creates dependencies among communication channels, hence indirectly among join patterns. In [33], we have presented JEScala, a language that captures coordination schemas in a more expressive and modular way by leveraging a seamless integration of an advanced event system with join abstractions. We have validated the approach with case studies and provided a first performance assessment.

*6.2.2.2. Lazy imperative programming*

Laziness is a powerful concept in functional programming that permits the reuse of general functions in a specific context, while keeping performance close to the efficiency of dedicated definitions. Lazy evaluation can be used in imperative programming too. Twenty years ago, John Launchbury was already advocating for lazy imperative programming, but the level of laziness of his framework remained limited. Twenty years after, the picture has not changed.

We have proposed an Haskell framework to specify computational effects of imperative programs as well as their dependencies [23]. We have presented a semantics of a call-by-need lambda-calculus extended with imperative strict and lazy features and proved the correctness of our approach. While originally motivated by a less rigid use of foreign functions, we have shown that our approach is fruitful for a simple scenario based on sorted mutable arrays. Furthermore, we can take advantage of equations between algebraic operations to dynamically optimize compositions of imperative computations.

*6.2.2.3. Domain-Specific Aspect Languages*

Domain-Specific Aspect Languages (DSALs) are Domain-Specific Languages (DSLs) designed to express crosscutting concerns. Compared to DSLs, their aspectual nature greatly amplifies the language design space. In the context of the Associate Team RAPIDS/REAL, we have structured this space in order to shed light on and compare the different domain-specific approaches to deal with crosscutting concerns [37]. We have reported on a corpus of 36 DSALs covering the space, discussed a set of design considerations and provided a taxonomy of DSAL implementation approaches. This work serves as a frame of reference to DSAL and DSL researchers, enabling further advances in the field, and to developers as a guide for DSAL implementations.

*6.2.2.4. Controlling constraint propagation*

Constraint propagation is at the heart of constraint solvers. Two main trends co-exist for its implementation: variable-oriented propagation engines and constraint-oriented propagation engines. These two approaches ensure the same level of local consistency but their efficiency (computation time) can be quite different depending on the problem instances to be solved. However, it is usually accepted that there is no best approach in general, and modern constraint solvers implement only one of them.

In the context of Charles Prud'homme's PhD Thesis [15], we have gone a step further providing a solver independent language at the modeling stage to enable the design of propagation engines. We have validated our proposal with a reference implementation based on the Choco solver and the MiniZinc constraint modeling language.

## 6.3. Software Composition

**Participants:** Diana Allam, Walid Benghabrit, Ronan-Alexandre Cherrueau, Rémi Douence, Hervé Grall, Thomas Ledoux, Jean-Claude Royer, Mohamed Sellami, Mario Südholt.

### *6.3.1. Constructive Security*

Nowadays we are witnessing the wide-spread use of cloud services. As a result, more and more end-users (individuals and businesses) are using these services for achieving their electronic transactions (shopping, administrative procedures, B2B transactions, etc.). In such scenarios, personal data is generally flowing between several entities and end-users need (i) to be aware of the management, processing, storage and retention of personal data, and (ii) to have necessary means to hold service providers accountable for the usage of their data. Usual preventive security mechanisms are not adequate in a world where personal data can be exchanged on-line between different parties and/or stored at multiple jurisdictions. Accountability becomes a necessary principle for the trustworthiness of open computer systems. It regards the responsibility and liability for the data handling performed by a computer system on behalf of an organization. In case of misconduct (e.g. security breaches, personal data leak, etc.), accountability should imply remediation and redress actions, as in the real life.

In 2014, we have developed two general approaches for the definition and enforcement of accountability properties.

*6.3.1.1. Logic-based accountability properties*

We have proposed a framework for the representation of cloud accountability policies [19]. Such policies offer end-users a clear view of the privacy and accountability obligations asserted by the entities they interact with, as well as means to represent their preferences. This framework comes with two novel accountability policy languages; an abstract one, which is devoted for the representation of preferences/obligations in an human

readable fashion, a concrete one for the mapping to concrete enforceable policies. We motivate our solution with concrete use case scenarios. [30] discusses issues related to data privacy and big data technologies and advocate the use of the framework to support accountability.

We have provided an abstract language for the representation of accountability obligations [20]. We define its semantics using first-order temporal logic and a specific modality for accountability is introduced. We analyze a healthcare use case to illustrate the efficiency of our approach in representing accountability obligations in realistic situations. The use of such services-based applications usually implies the flow of personal data on-line between several parties. In [21], we consider this issue at the design-time of the software and we propose some foundations for an accountable software design. Accountability for a software is a property describing, among other aspects, its liability to end-users for the usage of the data it has been entrusted. We propose to enrich software's component design by accountability clauses using an abstract accountability language (introduced in [20]). We also define conditions for the well-formedness of an accountable component design and show how they can be checked using the $\mu$-CRL model-checker.

*6.3.1.2. Defining and enforcing multi-level accountability properties*

Many accountability policies require access to all levels of the software stack of service-based applications. Furthermore, they should include explicit means for the definition of cross-domain policies and provide constructive means for the implementation of a wide variety of of accountability properties. These features, in particular, multi-level support, are missing in existing approaches.

We have provided an approach that addresses these objectives explicitly through a language for the definition of expressive regular policies over accountability predicates applicable at all levels of the service stack [22]. Furthermore, we have presented hierarchies of constructive schemes for the implementation of policies for transparency and remediation properties that are implemented in terms of our accountability policy language. Finally, we have shown how to harness the accountability schemes to tackle real-world violations of accountability properties arising from security vulnerabilities of OAuth-based authorization and authentication protocols.

### *6.3.2. Aspect-Oriented Programming*

We have produced in 2014 a range of results enabling reasoning over aspect languages and investigated the use of execution levels. These results are presented in the remainder of this section.

We have also applied ideas from aspect oriented programming in the context of distributed programming (aspectual session types [32]), see Sec. 6.4 .

*6.3.2.1. Reasoning about aspect interference using effective aspects*

Aspect-oriented programming (AOP) aims at enhancing modularity and reusability in software systems by offering an abstraction mechanism to deal with crosscutting concerns. But, in most general-purpose aspect languages aspects have almost unrestricted power, eventually conflicting with these goals. To tame aspects, we have proposed Effective Aspects: a novel approach to embed the pointcut/advice model of AOP in a statically-typed functional programming language like Haskell; along two main contributions. First, we have defined a monadic embedding of the full pointcut/advice model of AOP [16].

Type soundness is guaranteed by exploiting the underlying type system, in particular phantom types and a new anti-unification type class. In this model aspects are first-class, can be deployed dynamically, and the pointcut language is extensible, therefore combining the flexibility of dynamically-typed aspect languages with the guarantees of a static type system. Monads (which allow the definition of sequences of computations in functional programs) enable us to directly reason about computational effects both in aspects and base programs using traditional monadic techniques. Using this we extend the notion of Open Modules with effects, and also with protected pointcut interfaces to external advising. These restrictions are enforced statically using the type system. Also, we adapt the techniques of EffectiveAdvice to reason about and enforce control flow properties as well as to control effect interference. We show that the parametricity-based approach to effect interference falls short in the presence of multiple aspects and propose a different approach using monad views, a novel technique for handling the monad stack, developed by Schrijvers and Oliveira. Then, we

exploit the properties of our model to enable the modular construction of new semantics for aspect scoping and weaving. Our second contribution [24] builds upon a powerful model to reason about mixin-based composition of effectful components and their interference, based on equational reasoning, parametricity, and algebraic laws about monadic effects. Our contribution is to show how to reason about interference in the presence of unrestricted quantification through pointcuts. We show that global reasoning can be compositional, which is key for the scalability of the approach in the face of large and evolving systems. A comprehensive version of those two works appears in Ismael Figueroa PhD thesis [12].

*6.3.2.2. Execution Levels for AOP: from program design to applications*

In AOP languages, advice evaluation is usually considered as part of the base program evaluation. This is also the case for certain pointcuts, such as if pointcuts in AspectJ, or simply all pointcuts in higher-order aspect languages like AspectScheme. While viewing aspects as part of base level computation clearly distinguishes AOP from reflection, it also comes at a price: because aspects observe base level computation, evaluating pointcuts and advice at the base level can trigger infinite regression. To avoid these pitfalls, aspect languages propose ad hoc mechanisms, which increase the complexity for programmers while being insufficient in many cases. We have proposed to clarify the situation by introducing levels of execution in the programming language [18], thereby allowing aspects to observe and run at specific, possibly different, levels. We have adopted a defensive default that avoids infinite regression, and gives advanced programmers the means to override this default using level-shifting operators.

### *6.3.3. Service provisioning*

This year, we have provided results on two fundamental problems of service-oriented architectures: service interoperability and service mediation.

*6.3.3.1. Service interoperability*

Web service support a document-oriented style for clients to interact with a server and promote an environment for systems that is loosely coupled and interoperable. Two models exist for implementing Web services: A process-oriented Web services model, SOAP, and a resource-oriented Web services model, RESTful. Service components are mainly based on description interfaces. These interfaces are often known as structural standardized interfaces like WSDL for SOAP and WADL for RESTful. The implementation of Web services is increasingly based on object-oriented (OO) frameworks, at the client and the server sides. Using these frameworks, developers can transform an object code into a Web service, or access a remote Web service, at the touch of a button. In this context, two levels are present: an object level built over a service level.

Diana Allam's PhD thesis [11] has focused on two properties of these frameworks:

- The loose coupling between the two levels, which allows the complex technical details of the service level to be hidden at the object level and the service level to be evolved with a minimal impact on the object level.
- The interoperability induced by the substitution principle associated to subtyping in the object level, which allows to freely convert a value of a subtype into a value of a supertype.

The thesis provides three contributions in this context. We propose a unified formal model for web services based on message passing and enabling first class channels. It is equipped with a powerful type-checking allowing union, intersection and negation operations as well as subtyping. The type checking algorithm relies on the semantic approach defined by G. Castagna. This type system is also protected against attackers. The second contribution is a concrete refinement of the model into RESTful and SOAP frameworks as well as a unified API for service discovery. To define such an API, we have first shown how the details of the standard interfaces (WSDL and WADL) could be simplified and abstracted and then we rely on subtyping in the discovery mechanism. Finally, to solve some of the interoperability issues between the OO level and the service level a formalization of the binding using categorical concepts (commutative diagrams) is proposed. Based on this an analysis of the mismatch problems has been done and a new specification of the data binding has been formalized. The document then discusses some variations in the implementation of the data binding solution and a prototype for the Apache CXF framework.

Mayleen Lacouture's PhD thesis "A Chemical Programming Language for Orchestrating Services - Application to Interoperability Problems" [M. Lacouture, MN/U. Nantes, Oct. 14] proposes a framework easing interoperability in the form of an architecture that integrates different orchestration languages with heterogeneous service providers around a pivot language. The pivot language is implemented as a new orchestration language based on the chemical programming paradigm. Concretely, the dissertation presents a language called Criojo that implements and extends the Heta-calculus, an original calculus associated to a chemical abstract machine dedicated to service-oriented computing. The consequence of adopting this approach would be an improvement in the interoperability of services and orchestration languages, thus easing the development of composite services. The high level of abstraction of Criojo could allow developers to write very concise orchestrations since message exchanges are represented in a natural and intuitive way.

*6.3.3.2. Service mediation*

Service composition is a major advance service-oriented computing brings to enable the development of distributed applications. However, the distributed nature of services hampers their composition with data heterogeneity problems. We address these problems with a decentralized Mediation-as-a-Service architecture that solves data inconsistencies occurring during the composition of business services [17]. As an extension to our previous work that focused on data interpretation problems, we present in this paper a solution to solve data inconsistencies at the syntactic, structural and semantic levels. We show how syntactic, structural and semantic mediation techniques can be combined, and how semantic mediation provides useful information that helps structural and syntactic mediation. We demonstrate how our architecture enables decentralized publication and discovery of mediation services. We motivate our work with a concrete scenario and validate our proposal with experiments.

### *6.3.4. Software product line architectures*

Software product lines were designed from the product line tested out by H. Ford at the beginning of the 20TH century, which led to the success of his automotive production. For 15 years, these methods have been visible in several software application fields: telephony at Nokia, televisions at Philips, print software at HP and flight applications at Boeing, among others. The concept of architecture is crucial for classic software applications, and this concept is even more important at the level of domain engineering in product lines. In a product line, the so-called reference architecture generically describes the architectures of all the products in the family. The chapter [34] describes the technical means and methods for defining a reference architecture for a software product line. It also presents the methods for operating this architecture through, for example, techniques emerging from model and software component engineering, or aspect-oriented programming. These concepts and techniques are illustrated using a case study.

## 6.4. Cloud applications and infrastructures

**Participants:** Adrien Lebre, Thomas Ledoux, Yousri Kouki, Guillaume Le Louët, Jean-Marc Menaud, Jonathan Pastor, Flavien Quesnel, Mario Südholt.

In 2014, we have provided solutions for Cloud-based and distributed programming, virtual environments and data centers, in particular concerning energy-optimal Cloud applications.

### *6.4.1. Cloud and distributed programming*

This year we have published results on a broker that provides better guarantees on service-level agreements in the Cloud. Furthermore, we have extended a class of formally-defined protocols, session types.

*6.4.1.1. Service-level agreement for the Cloud*

Elasticity is the intrinsic element that differentiates Cloud Computing from traditional computing paradigms, since it allows service providers to rapidly adjust their needs for resources to absorb the demand and hence guarantee a minimum level of Quality of Service (QoS) that respects the Service Level Agreements (SLAs) previously defined with their clients. However, due to non-negligible resource initiation time, network fluctuations or unpredictable workload, it becomes hard to guarantee QoS levels and SLA violations may occur.

We propose a language support for Cloud elasticity management that relies on CSLA (Cloud Service Level Agreement) [27]. CSLA offers new features such as QoS/functionality degradation and an advanced penalty model that allow providers to finely express contracts so that services self-adaptation capabilities are improved and SLA violations minimized. The approach was evaluated with a real infrastructure and application testbed. Experimental results show that the use of CSLA makes Cloud services capable of absorbing more peaks and oscillations by trading-off the QoS levels and costs due to penalties.

*6.4.1.2. AO session types for distributed protocols*

Multiparty session types allow the definition of distributed processes with strong communication safety properties. A global type is a choreographic specification of the interactions between peers, which is then projected locally in each peer. Well-typed processes behave accordingly to the global protocol specification. Multiparty session types are however monolithic entities that are not amenable to modular extensions. Also, session types impose conservative requirements to prevent any race condition, which prohibit the uniform application of extensions at different points in a protocol. We have proposed a means to support modular extensions with aspectual session types [32], a static pointcut/advice mechanism at the session type level. To support the modular definition of crosscutting concerns, we have augmented the expressivity of session types to allow harmless race conditions. As a result, aspectual session types make multiparty session types more flexible, modular, and extensible.

### *6.4.2. Virtualization and data centers*

In 2014, we have produced a variety of results on a new model for utility computing that addresses fundamental shortcomings of today's Cloud computing model. Furthermore, we have provided more powerful techniques for the virtualization of computations and the management of cluster-based environments, such as data centers.

*6.4.2.1. Next generation utility computing*

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 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. Throughout the Discovery initiative [0], we investigate how UC resources can be managed differently, considering locality as a primary concern. Concretely, we study how it can be possible to leverage any facilities available through the Internet in order to deliver widely distributed UC platforms that can better match the geographical dispersal of users as well as the unending resource demand. Critical to the emergence of such locality-based UC (LUC) platforms is the availability of appropriate operating mechanisms. We presented a prospective vision of a unified system driving the use of resources at an unprecedented scale by turning a complex and diverse infra structure into a collection of abstracted computing facilities that is both easy to operate and reliable [35]. By deploying and using such a LUC Operating System on backbones, our ultimate vision is to make possible to host/operate a large part of the Internet by its internal structure itself: A scalable and nearly infinite 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. We highlight that this work is conducted through a collaboration between the ASAP, ASCOLA, AVALON and MYRIADS Inria Project-teams.

*6.4.2.2. Adding locality capabilities to virtual machine schedulers*

Through the DVMS proposal, we showed in 2013 the benefit of leveraging peer-to-peer algorithms to design and implement virtual machines (VMs) scheduling algorithms. Although P2P based proposals considerably improve the scalability, leading to the management of hundreds of thousands of VMs over thousands of physical machines (PMs), they do not consider the network overhead introduced by multi-site infrastructures. This over- head can have a dramatic impact on the performance if there is no mechanism favoring intra-site v.s. inter-site manipulations. This year, we extended our DVMS mechanism with a new building block designed on top of the Vivaldi coordinates mechanism. We showed its benefits by discussing several experiments performed

[0] http://beyondtheclouds.github.io

on four distinct sites of the Grid'5000 testbed. With our proposal and without changing the scheduling decision algorithm, the number of inter-site operations has been reduced by 72% [29]. This result provides a glimpse of the promising future of using locality properties to improve the performance of massive distributed Cloud platforms. We highlight that this work has been performed in collaboration with the ASAP, ASCOLA, AVALON and MYRIADS Inria Project-teams.

*6.4.2.3. WAN-wide elasticity capabilities for distributed file systems*

Applications dealing with huge amounts of data suffer significant performance impacts when they are deployed on top of an hybrid platform (i.e the extension of a local infrastructure with external cloud resources). More precisely, through a set of preliminary experiments we shew that mechanisms which enable on demand extensions of current Distributed File Systems (DFSes) are required. These mechanisms should be able to leverage external storage resources while taking into account the performance constraints imposed by the physical network topology used to interconnect the different sites. To address such a challenge we presented the premises of the Group Based File System, a glue providing the elasticity capability for storage resources by federating on demand any POSIX file systems [28].

### *6.4.3. Energy optimization*

Demand for Green services is increasing considerably as people are getting more environmental conscious to build a sustainable society. Therefore, enterprise and clients want to shift their workloads towards green Cloud environment offered by the Infrastructure-as-a-Service (IaaS) provider. The main challenge for an IaaS provider is to determine the best trade-off between its profit while using renewable energy and customers satisfaction. In order to address this issue, we propose a *Cloud energy broker* [26], which can adjust the availability and price combination to buy Green energy dynamically from the market to make datacenter green. Our energy broker tries to maximize of using renewable energy under strict budget constraint whereas it also tries to minimize the use of brown energy by capping the limit of overall energy consumption of datacenter. The energy broker was evaluated with a real workload traced by PlanetLab. Experimental results show that our energy broker successfully enables meeting the best trade-off.

**ASPI Project-Team**

# 5. New Results

## 5.1. Adaptive multilevel splitting

**Participants:** Frédéric Cérou, Arnaud Guyader.

We show [21] that an adaptive version of multilevel splitting for rare events is strongly consistent. We also show that the estimates satisfy a CLT (central limit theorem), with the same asymptotic variance as the non-adaptive algorithm with the optimal choice of the parameters. It is a strong and general result, that generalizes some of our previous results, and the proof is quite technical and involved.

This work has been presented at the 10th International Workshop on Rare Event Simulation (RESIM), held in Amsterdam in August 2014.

## 5.2. Convergence of a two–step multilevel splitting algorithm for rare event simulation

**Participants:** François Le Gland, Damien–Barthélémy Jacquemart.

The problem is to accurately estimate the (very small) probability that a rare but critical event (such as a score function exceeding a given threshold) occurs before some fixed final time. Multilevel splitting is a very efficient solution, in which sample paths are propagated and are eliminated or replicated when some intermediate events (defined by some intermediate thresholds) occur. A common and efficient design is to define the next intermediate level as an empirical quantile of the running maximum of the score function along a surviving trajectory. However, it is practically impossible to remember when (at which time instant) and where (in which state) did each successful trajectory cross the empirically defined threshold. The proposed design is a two–step adaptive multilevel splitting algorithm: In the first step, a first set of trajectories is sampled in order to obtain the next intermediate threshold as an empirical quantile. In the second step, once the new intermediate threshold is obtained, a second set of trajectories is sampled in order to evaluate the transition probability to the new empirically defined intermediate region. This two–step procedure is repeated until some trajectories do hit the critical region before final time.

This work has been presented at the 10th International Workshop on Rare Event Simulation (RESIM), held in Amsterdam in August 2014.

## 5.3. Simulation–based algorithms for the optimization of sensor deployment

**Participant:** François Le Gland.

This is a collaboration with Christian Musso (ONERA, Palaiseau) and with Sébastien Paris (LSIS, université du Sud Toulon Var), related with the supervision of the PhD thesis of Yannick Kenné.

The problem considered here can be described as follows: a limited number of sensors should be deployed by a carrier in a given area, and should be activated at a limited number of time instants within a given time period, so as to maximize the probability of detecting a target (present in the given area during the given time period). There is an information dissymmetry in the problem: if the target is sufficiently close to a sensor position when it is activated, then the target can learn about the presence and exact position of the sensor, and can temporarily modify its trajectory so as to escape away before it is detected. This is referred to as the target intelligence. Two different simulation–based algorithms have been designed to solve separately or jointly this optimization problem, with different and complementary features. One is fast, and sequential: it proceeds by running a population of targets and by dropping and activating a new sensor (or re–activating a sensor already available) where and when this action seems appropriate. The other is slow, iterative, and non–sequential; it proceeds by updating a population of deployment plans with guaranteed and increasing criterion value at each

iteration, and for each given deployment plan, there is a population of targets running to evaluate the criterion. Finally, the two algorithms can cooperate in many different ways, to try and get the best of both approaches. A simple and efficient way is to use the deployment plans provided by the sequential algorithm as the initial population for the iterative algorithm.

This work has been presented at the Conference on Optimization and Practices in Industry (COPI), held in Palaiseau in October 2014.

## 5.4. Non–homogeneous Markov switching auto–regressive models for wind time series

**Participants:** Valérie Monbet, Julie Bessac.

This is a collaboration with Pierre Ailliot (UBO) and Françoise Pène (UBO).

We proposed [20] non–homogeneous Markov switching auto–regressive models for bivariate wind time series considering Cartesian coordinates on one hand and polar coordinates on the other hand. In non–homogeneous models, the transitions depend on the wind direction at the previous time. At the location of interest, wind is rotating more often clockwise but wind direction may also oscillate around two prevailing directions (northeast for anti–cyclonic conditions and southwest for cyclonic conditions). These features induce respectively some cycles which can be seen in the second order structure and modes in the marginal distribution. In broad outline, non–homogeneous transitions help the process to stay in the same weather regime when the wind direction is close to the prevailing directions and lead to sojourn duration in the regimes which are not geometric.

## 5.5. Gaussian state–space models for wind speed

**Participants:** Valérie Monbet, Julie Bessac.

This is a collaboration with Pierre Ailliot (UBO).

A multi–site stochastic generator for wind speed has been developped [11]. It aims at simulating realistic wind conditions with a focus on reproducing the space-time motions of the meteorological systems. A Gaussian linear state–space model is used where the latent state may be interpreted as regional wind conditions and the observation equation links regional and local scales. The model is fitted to 6–hourly reanalysis data in the North–East Atlantic. It is shown that it is interpretable and provides a good description of important properties of the space–time covariance function of the data, such as the non full–symmetry induced by prevailing flows in this area.

## 5.6. Level–dependent time deformation of Gaussian processes

**Participant:** Valérie Monbet.

Many records in environmental science exhibit asymmetries. In this project, we introduce a time deformation to produce asymmetric path from a Gaussian process with symmetric path. A simple case is obtained by assuming that

$$Z_t = Y_{\phi(t)}, \quad \phi(t) = \int_0^t f(Z_s)ds$$

with $\{Y_t\}$ a stationary Gaussian process. The function $f$ which controls the time deformation is increasing. The time–change function $\phi$ is such that the modified time increases quicker when the process is at high levels and thus that the crests of the modified process $\{Z_t\}$ are narrower than the ones of $\{Y_t\}$. The opposite holds true for the troughs. Inference tools are developed to estimate the function $f$.

## 5.7. Self–similar prior and wavelet bases for hidden turbulent motion

**Participant:** Patrick Héas.

This is a collaboration with Frédéric Lavancier (université de Nantes) and Souleymane Kadri–Harouna (université de la Rochelle)

This work [14] is concerned with the ill–posed inverse problem of estimating turbulent flows from the observation of an image sequence. From a Bayesian perspective, a divergence–free isotropic fractional Brownian motion (fBm) is chosen as a prior model for instantaneous turbulent velocity fields. This self–similar prior characterizes accurately second–order statistics of velocity fields in incompressible isotropic turbulence. Nevertheless, the associated maximum a posteriori involves a fractional Laplacian operator which is delicate to implement in practice. To deal with this issue, we propose to decompose the divergence–free fBm on well–chosen wavelet bases. As a first alternative, we propose to design wavelets as whitening filters. We show that these filters are fractional Laplacian wavelets composed with the Leray projector. As a second alternative, we use a divergence–free wavelet basis, which takes implicitly into account the incompressibility constraint arising from physics. Although the latter decomposition involves correlated wavelet coefficients, we are able to handle this dependence in practice. Based on these two wavelet decompositions, we finally provide effective and efficient algorithms to approach the maximum a posteriori.

## 5.8. Estimation of non–linear dynamics under sparse constraints

**Participant:** Patrick Héas.

This is a collaboration with Cédéric Herzet (EPI FLUMINANCE, Inria Rennes–Bretagne Atlantique) and Angélique Drémeau (ENSTA Bretagne, Brest).

Following recent contributions in non–linear sparse representations, this work [19], [18] focuses on a particular non–linear model, defined as the nested composition of functions. This family includes in particular discrete–time hidden Markov models. Recalling that most linear sparse representation algorithms can be straightforwardly extended to non–linear models, we emphasize that their performance highly relies on an efficient computation of the gradient of the objective function. In the particular case of interest, we propose to resort to a well–known technique from the theory of optimal control to evaluate the gradient. This computation is then implemented into the $\ell_1$–reweighted procedure proposed by Candès et al. [24], leading to a non–linear extension of it. As an example, we consider the problem of estimating the ocean state from satellite low–dimensional information by exploiting a geophysical dynamical model and a sparse decomposition of the initial condition in some redundant dictionary.

This work has also been presented at Congrès National d'Assimilation, a national event held in Toulouse in December 2014.

**ATLANMOD Project-Team**

# 6. New Results

## 6.1. Model Quality

Our work aims to enhance the quality of the modeling activity in the context of software engineering and language engineering. This year, this has translated in the following results:

- A systematic review [16] of all formal verification approaches targeting the quality evaluation of software models to be used as the basis for future research on the topic and as a kind of reference comparison to compare new tools with existing ones.
- A complete description of our CSP-based approach for the verification of UML/OCL models (where both the uml constructs and OCL expressions are translated into a constraint satisfaction problem) [12]
- A new test data generation approach for Model Transformations that combines partitions and constraint analysis to try to mazimize the coverage of the generates tests [29]

## 6.2. Model Driven approach to mobile applications development

Cross-platform and multi-device design, implementation and deployment is a barrier for today's IT solution providers, especially SME providers, due to the high cost and technical complexity of targeting development to a wide spectrum of devices, which differ in format, interaction paradigm, and software architecture. Our work aims at exploiting the modern paradigm of Model-Driven Engineering and code generation to simplify multi-device development, reducing cost and development times, so as to increase the profit of SME solution providers and at the same time reduce the price and total cost of ownership for end-customers. In [22] we defined a Platform Independent Modeling language for mobile applications. The language has been defined as a mobile extension of an OMG standard called Interaction Flow Modeling Language (IFML). The research included also the development of an Eclipse-based modeling tool for mobile apps and the first prototypes of automatic code generators.

## 6.3. Security

Most companies information systems are composed by heterogeneous components responsible of hosting, creating or manipulating critical information for the day-to-day operation of the company. Securing this information is therefore one of their main concerns, more particularly specifying Access Control (AC) policies. However, the task of implementing an AC security policy (sometimes relying on several mechanisms) remains complex and error prone as it requires knowing low level and vendor-specific facilities. In this context, discovering and understanding which security policies are actually being enforced by the Information System (IS) becomes critical. Thus, the main challenge consists in bridging the gap between the vendor-dependent security features and a higher-level representation. This representation has to express the policies by abstracting from the specificities of the system components, allowing security experts to better understand the policy and to implement all related evolution, refactoring and manipulation operations in a reusable way.

In 2014, we have presented a Ph.D. thesis tackling the aforementioned problems. It proposes a model-driven automatic reverse engineering mechanism capable of analyzing deployed security aspects of components (e.g. concrete firewall configurations) to derive the abstract model (e.g. network security global policy) that is actually enforced. Once the model is obtained, it can be reconciled with the expected security directives, to check its compliance, can be queried to test consistency or used in a process of forward engineering to generate validated security configurations. This work also provides the first steps towards the integration of the diverse security policies extracted from the subsystems composing a complex Information System in a global security representation.

## 6.4. Model-Driven Document Engineering

As a result of a long-term collaboration of one of the AtlanMod team members with the ISSI research group at the Universitat Politècnica de València, we have participated in the publication of several works on the area of the Document Engineering. In this research line, we have applied the MDE methods and tools to the product-line-based generation of customized documents resulting in the so-called DPL methodology [0]. The Document Product Lines (DPL) approach, which we throughly describe in a journal publication [17], provides a framework for variable content document generation that follows an alternate path to the traditional variable document generation. DPL has been created with a twofold goal: first, to make creating variable content documents available to non-experts by including a domain engineering process previous to the document generation itself; and secondly, to enforce content reuse at domain level.

DPLFW is the main tool supporting the DPL methodology, and in the demonstrations track of the MODELS conference we showed all its capabilities. In addition to these contributions, we have published several works demonstrating the applicability of the DPL–DPLFW tandem in different domains, such as the development of executable emergency plans in crisis managent contexts [25], the development of learning objects in the e-learning field [32] and the generation of customized documents in e-Government solutions [33].

## 6.5. Reverse Engineering and Evolution

Model Driven Reverse Engineering (MDRE), and its applications such as software modernization, is a discipline in which model-driven development (MDD) techniques are used to treat legacy systems. During this year, Atlanmod has continued working actively on this research area. The main contributions are the following:

- In the context of the ARTIST FP7 project, the work has been continued on reusing (and extending accordingly) MoDisco and several of its components to provide the Reverse Engineering support required within the project. At conceptual-level, the MoDisco Model Discovery + Model Understanding overall two-step approach [11] has been published and promoted as an important part of the ARTIST migration methodology and process [18]. At tooling-level, several (MoDisco-based) model discovery components from Java and SQL have been developed and made available as part of the official ARTIST OS Release [0]. Directly related to some of these components, a promising work has been initiated on studying deeper the discovery of behavioral aspects of software and dealing with their further understanding based on the OMG FUML standard combined with different modeling techniques (transformation, slicing, etc.). Complementary work has also been performed in the context of the TEAP FUI project finishing by the end of this year. It concerns the related problem of data federation from heterogeneous sources in the domain of Enterprise Architecture. This has notably resulted in a prototype called EMF Views that can be practically used in such reverse engineering scenarios [36] and also in other cases to be further explored (cf. the MoNoGe FUI project dealing with (meta)model extension).
- In a web context, in a previous work we shown how to discover the schema which is implicit in JSON data. This year we built on that contribution to study how schemas coming from different JSON-based web APIs can be composed [24]. Thus, we presented an approach able to identify composition links between schemas of different APIs. This composition information plus the API schemas are used to render a graph where paths represent API compositions and are used to easily identify how to compose the APIs. For instance, we illustrated one application based on generating sequence diagrams from graph paths, where the diagram includes the API calls (and their corresponding parameters) that web developers have to perform in order to compose one or more APIs.
- In the context of our work around DSLs, we have been working on facilitating the definition of DSLs from existing APIs. Sometimes library developers prefer to provide their users with a DSL, instead of (or in addition to) an API. APIs and DSLs can be seen as alternative methods to access the library

[0] http://dpl.dsic.upv.es, only in Spanish
[0] http://www.artist-project.eu/tools-of-toolbox/193

functionalities, and are characterized by specific advantages. We therefore proposed a method to automatically analyze an existing object-oriented API and generate a DSL out of it. Our approach leverages on model-driven techniques to analyze and represent APIs at high-level of abstraction (i.e., as metamodels) which are later used to automatically generate the DSL components and the corresponding tooling, including parser, compiler and development environment. Developers can influence the DSL generation by editing the model-based API representation and by specifying design choices about the structure of the DSL to generate. A proof-of-concept implementation of the method has been developed, called *DSLit*, that is able to analyze Java APIs and generate textual DSLs.

- On the evolution side, we have been working on an approach to automatically resynchronize code-generation artefacts (in particular, model-to-text transformations) after changes on the target platform [28]

## 6.6. Scalability

The increasing number of companies embracing MDE methods and tools have exceeded the limits of the current model-based technologies, presenting scalability issues while facing the growing complexity of their data. Since further research and development is imperative in order to maintain MDE techniques as relevant as they are in less complex contexts, we have focused our research in three axes, (i) scalable persistence solutions, (ii) scalable model transformation engines, and (iii) testing of large scale distributed systems.

In [21], we lead the first open-set benchmark gathered from real-world cases to stress scalability issues in model transformation and query engines. This benchmark suite has been made public with a twofold goal: (i) to provide a reference benchmark suite to both the industry and the research community that can be used to compare and evaluate different technologies that may fulfill their needs; and (ii) to motivate the MDE community to be part of its development by allowing them to extend and contribute with additional cases not covered by the initial set.

On the other hand, we introduce Neo4EMF [20], a NoSQL database persistence framework based on Neo4j [0]. Neo4EMF provides light-weight on-demand loading and storage facilities for handling very large models. Additionally, we also show that Neo4EMF can handle the creation of very-large models without performing periodical saves manually.

In this paper [31], we argue that fUML may be leveraged to address the well-known interoperability issue between tools from different modeling platforms. This is done by providing a common execution language and by abstracting modeling frameworks into generic actions that perform elementary operations on models. User models can not only benefit from a unified execution semantics, but also modeling tools can benefit too. As a proof of concept, we show [37] how it can be applied to model transformation engines, in particular ATL. To this end, an prototype compiler from ATL to fUML has been built.

In [19], we present a model-based approach to define a dynamic oracle for checking global properties on distributed software. Our objective is to abstract relevant aspects of such systems into models by gathering data from different nodes and building a global view of the system, where properties are validated. These models are updated at runtime, by monitoring the corresponding distributed system. This process requires a distributed test architecture and tools for representing and validating global properties. To evaluate the ability of our approach, a real-scale experimental validation has been conducted.

[0] http://www.neo4j.org

**CAIRN Project-Team**

# 6. New Results

## 6.1. Highlights of the Year

Our work on accuracy evaluation and optimisation for fixed point arithmetic was presented during a tutorial "Automatic Fixed-Point Conversion: a Gateway to High-Level Power Optimization" at IEEE/ACM Design Automation and Test in Europe [77].
As a proof of concept of our research on improving efficiency of dynamic reconfiguration in FPGAs [47] [48], the *eFPGA* (Figure 5 ) chip was designed and fabricated in 65nm CMOS technology. In the proposed and patented architecture [73] (EU patent), the configuration of the FPGA becomes independent from its placement and is moreover significantly compressed (up to $\times 10$). This notion of *Virtual Bit Stream* allows for seamless partial and dynamic reconfiguration and for task migration.

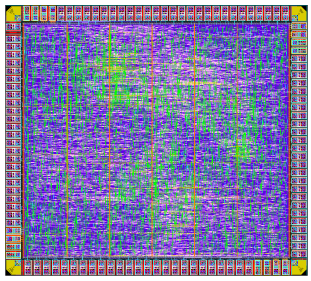

*Figure 5.* CAIRN*'s eFPGA chip*

## 6.2. Reconfigurable Architecture Design

### *6.2.1. Dynamic reconfiguration support in FPGA*

**Participants:** Olivier Sentieys, Antoine Courtay, Christophe Huriaux.

Almost since the creation of the first SRAM-based FPGAs there has been a desire to explore the benefits of partially reconfiguring a portion of an FPGA at run-time while the remainder of design functionality continues to operate uninterrupted. Currently, the use of partial reconfiguration imposes significant limitations on the FPGA design: reconfiguration regions must be constrained to certain shapes and sizes and, in many cases, bitstreams must be precompiled before application execution depending on the precise region of the placement in the fabric. We plan to develop an FPGA architecture that allows for seamless translation of partially-reconfigurable regions, even if the relative placement of fixed-function blocks within the region is changed.

**FPGA Architecture Support for Heterogeneous, Relocatable Partial Bitstreams.**

The use of partial dynamic reconfiguration in FPGA-based systems has grown in recent years as the spectrum of applications which use this feature has increased. For these systems, it is desirable to create a series of partial bitstreams which represent tasks which can be located in multiple regions in the FPGA fabric. While the transferal of homogeneous collections of lookup-table based logic blocks from region to region has been shown to be relatively straightforward, it is more difficult to transfer partial bitstreams which contain fixed-function resources, such as block RAMs and DSP blocks. In this work we consider FPGA architecture enhancements which allow for the migration of partial bitstreams including fixed-function resources from region to region even if these resources are not located in the same position in each region. Our approach does not require significant, time-consuming place-and-route during the migration process. We quantify the cost of inserting additional routing resources into the FPGA architecture to allow for easy migration of heterogeneous, fixed-function resources. Our experiments show that this flexibility can be added for a relatively low overhead and performance penalty. This work was performed during Christophe Huriaux's visit at UMASS in summer 2014 in the context of Inria Associate Team Hardiesse and has been published in [48] and in [74] as a poster.

**Virtual Bit Streams: Design Flow and Run-Time Management of Compressed and Relocatable FPGA Configurations.**

The aim of partially and dynamically reconfigurable hardware is to provide an increased flexibility through the load of multiple applications on the same reconfigurable fabric at the same time. However, a configuration bit-stream loaded at runtime should be created offline for each task of the application. Moreover, modern applications use a lot of specialized hardware blocks to perform complex operations, which tends to cancel the "single bit-stream for a single application" paradigm, as the logic content for different locations of the reconfigurable fabric may be different. We proposed a design flow for generating compressed configuration bit-streams abstracted from their final position on the logic fabric. Those configurations can then be decoded and finalized in real-time and at run-time by a dedicated reconfiguration controller to be placed at a given physical location. The VTR framework has been expanded to include bit-stream generation features. A bit-stream format is proposed to take part of our approach and the associated decoding architecture was designed. We analyzed the compression induced by our coding method and proved that compression ratios of at least $2.5\times$ can be achieved on the 20 largest MCNC benchmarks. The introduction of clustering which aggregates multiple routing resources together showed compression ratio up to a factor of $10\times$, at the cost of a more complex decoding step at runtime. Future perspectives on the VBS include extension of the architecture to support commercially available FPGAs as well as the improvement of the associated CAD tool flow to include smarter coding of the VBS to gain in runtime efficiency and in size. The VBS approach can provide increased online relocation capabilities using a decoding algorithm capable of decoding the VBS on-the-fly during the task migration. We applied for a European Patent on this work [73] and the results will be published in 2015 at IEEE/ACM DATE [47].

### *6.2.2. Power Models of Reconfigurable Architectures*

**Participants:** Robin Bonamy, Daniel Chillet, Olivier Sentieys.

Including a reconfigurable area in complex systems-on-chip is considered as an interesting solution to reduce the area of the global system and to support high performance. But the key challenge in the context of embedded systems is currently the power budget and the designer needs some early estimations of the power consumption of its system. Power estimation for reconfigurable systems is a difficult issue since several parameters need to be taken into account to define an accurate model. In this research, we consider the opportunity of the dynamic reconfiguration for the reduction of power consumption by the management of tasks scheduling and placement. We analyzed the power consumption during the dynamic reconfiguration on a Virtex 5 board. Three models of the partial and dynamic reconfiguration power consumption with different complexity/accuracy tradeoffs are defined. These models are used in design space exploration to evaluate the impact of reconfiguration on energy consumption of a complete system. We propose a methodology for power/energy consumption modeling and estimation in the context of heterogeneous (multi)processor(s) and dynamically reconfigurable hardware systems. We developed an algorithm to explore all task mapping possibilities for a complete application (e.g., for H264 video coding) with the aim to extract one of the best solutions with respect to the designer's requirements. This algorithm is a step ahead for defining on-line power

management strategies to decide which task instances must be executed to efficiently manage the available power using dynamic partial reconfiguration [24].

### 6.2.3. Real-time Spatio-Temporal Task Scheduling on 3D Architecture

**Participants:** Quang-Hai Khuat, Quang Hoa Le, Emmanuel Casseau, Antoine Courtay, Daniel Chillet.

One of the main advantages offered by a three-dimensional system-on-chip (3D SoC) is the reduction of wire length between different blocks of a system, thus improving circuit performance and alleviating power overheads of on-chip wiring. To fully exploit this advantage, an efficient management referring to allocate temporarily the tasks at different levels of the architecture is greatly important. In the context of 3D SoC, we have developed several spatio-temporal scheduling algorithms for 3D MultiProcessor Reconfigurable System-on-Chip (3DMPRSoC) architectures composed of a multiprocessor layer and an embedded Field Programmable Gate Array (eFPGA) layer with dynamic reconfiguration. These two layers are interconnected vertically by through-silicon vias (TSVs) ensuring tight coupling between software tasks on processors and associated hardware accelerators on the eFPGA. Our algorithms cope with task dependencies and try to allocate communicating tasks close to each other in order to reduce direct communication cost, thus reducing global communication cost. In the 3DMPRSoC context, our algorithms favor direct communications including: i) point-to-point communication between hardware accelerators on the eFPGA, ii) communication between software tasks through the Network-on-Chip of the multiprocessor layer, and iii) communication between software task and accelerator through TSV. When a direct communication between two tasks occurs, the data are stored in a shared memory placed onto the multiprocessor layer.

The algorithm proposed in [50] considers heterogenenous reconfigurable architecture and proposes a mathematical formulation for spatio-temporal scheduling of a task graph. The placement consists in finding the best mapping of the application task model onto the reconfigurable region. To improve the performance of our algorithm, we propose to configure the tasks by taking account of their priority. The global objective consists in the reduction of the global execution time. The second algorithm presented in [51] improves the previous one and proposes to exploit the presence of processor in the multiprocessor layer in order to anticipate a software execution of a task when no sufficient area is available. In this case, classical algorithms reject the task, and continue their execution. Our algorithm starts a software execution of the task, but the software execution is a speculative execution. Indeed, if a sufficient area is freed by a hardware task later, in this case our algorithm evaluates if the software execution must continue or if it is better to stop this execution to restart the task in the reconfigurable area. We demonstrated that the execution time of an application can be significantly reduced by applying this software speculation.

In [53], we proposed a heuristic which focus on the online task placement problem on a multi-context, dynamically and partially reconfigurable heterogeneous architecture. Configuration prefetching and anti-fragmentation well known techniques are combined with the place reservation technique that takes into account tasks to be placed in the future (pre-allocated tasks) while fulfilling task execution deadline constraint. Compared to a placement without reservation, our approach improves the number of placed tasks and the resource utilization rate.

### 6.2.4. Run-time Task Management to Increase Resource Utilisation for Concurrent Critical Tasks in Mixed-Critical Systems

**Participant:** Angeliki Kritikakou.

When integrating mixed critical systems on a multi/many-core system, one challenge is to ensure predictability for the high criticality tasks and an increased utilization for low criticality tasks. In [52], we proposed a distributed run-time WCET controller to address this problem, when several high criticality tasks with different deadlines, periods and offsets are concurrently executed on a multi core system.

During the system execution, the proposed controller regularly checks locally at each critical task if the interferences due to the low criticality tasks can be tolerated. This is achieved by monitoring the ongoing execution time, dynamically computing the remaining worst case execution time of the critical task when only critical tasks are executed on the system and checking our safety condition. In case that the condition is

violated for one critical task, the concurrent execution of the low criticality tasks with the critical one will lead to its deadline miss. Therefore, the local controller decides the suspension of the less critical tasks. However, the local controller is not responsible for the actual suspension of the low criticality tasks. The controller sends a request to a master which has a global view of the system. The master is in charge of collecting the requests of the critical tasks, suspending and restarting the low criticality tasks. When at least one critical task sends the request for suspension of the low criticality tasks, the master suspends them. During execution, the master updates the number of active requests and it restarts the low criticality tasks when all requesters have finished their execution. We have implemented our approach as a software controller on a real multi-core COTS system, the TMS320C6678 chip of Texas Instruments, where we have observed significant gains up to 556% for our case study.

### *6.2.5. Arithmetic Operators for Cryptography and Fault-Tolerance*

**Participants:** Arnaud Tisserand, Emmanuel Casseau, Nicolas Veyrat-Charvillon, Karim Bigou, Franck Bucheron, Jérémie Métairie, Gabriel Gallin, Huu Van Long Nguyen, Nicolas Estibals.

**Arithmetic Operators for Fast and Secure Cryptography.**

In the paper [39] presented at ASAP, we describe a new RNS (residue number system) modular multiplication algorithm, for finite field arithmetic over GF($p$), based on a reduced number of moduli in base extensions with only $3n/2$ moduli instead of $2n$ for standard ones. Our algorithm reduces both the number of elementary modular multiplications (EMMs) and the number of stored precomputations for large asymmetric cryptographic applications such as elliptic curve cryptography or Diffie-Hellman (DH) cryptosystem. It leads to faster operations and smaller circuits.

The PhD thesis defended by Karim Bigou [16] deals with the RNS representation and the associated arithmetic algorithms for asymmetric cryptography (ECC and RSA). The title of the PhD is "Theoretical Study and Hardware Implementation of Arithmetical Units in Residue Number System (RNS) for Elliptic Curve Cryptography".

Scalar recoding is popular to speed up ECC (elliptic curve cryptography) scalar multiplication: non-adjacent form, double-base number system, multi-base number system (MBNS). Ensuring uniform computation profiles is an efficient protection against some side channel attacks (SCA) in embedded systems. Typical ECC scalar multiplication methods use two point operations (addition and doubling) scheduled according to secret scalar digits. Euclidean addition chains (EAC) offer a natural SCA protection since only one point operation is used. Computing short EACs is considered as a very costly operation and no hardware implementation has been reported yet. We designed an hardware recoding unit for short EACs which works concurrently to scalar multiplication. It has been integrated in an in-house ECC processor on various FPGAs. The implementation results show similar computation times compared to non-protected solutions, and faster ones compared to typical protected solutions (e. g. 18 % speed-up over 192 b Montgomery ladder).

In the paper [40], we introduce a robust asynchronous logic family which does not rely on timing assumptions and/or delay elements and can operate with sub-powered devices. The key element behind our proposal is a simplified completion detection mechanism which makes it substantially more energy effective when compared with other dual-rail approaches. A 32-bit Ripple Carry Adder (RCA) is implemented in 65nm and 45nm CMOS process to evaluate the practicability of our approach. Firstly, the Optimal Energy Point (OEP) of the proposed RCA is investigated by scaling VDD from 0.4V to 0.2V (50mV interval), where the OEP occurs at 0.25V for both technologies. Secondly, while comparing the energy consumption with the corresponding single-rail benchmark at its OEP in 65nm process, 30% (34 fJ for 65nm) and 40% (54fJ for 45nm after scaling) energy savings are achieved respectively. More impressive (10x better) energy efficiency and reasonable performance are obtained over dual-rail counterparts. This work is done in the SPiNaCH project.

**ECC Crypto-Processor with Protections Against SCA.**

A dedicated processor for elliptic curve cryptography (ECC) is under development. Functional units for arithmetic operations in GF($2^m$) and GF($p$) finite fields and 160-600-bit operands have been developed for FPGA implementation. Several protection methods against side channel attacks (SCA) have been studied. The use of some number systems, especially very redundant ones, allows one to change the way some computations are performed and then their effects on side channel traces. This work is done in the PAVOIS project.

**Arithmetic Operators and Crypto-Processor for HECC.**

In the HAH project, we study and prototype efficient arithmetic algorithms for hyperelliptic curve cryptography for hardware implementations (on FPGA circuits). We study new advanced arithmetic algorithms and representations of numbers for efficient and secure implementations of HECC in hardware.

**Arithmetic Operators for Fault Tolerance.**

In the ARDyT and Reliasic projects, we work on computation algorithms, representations of numbers and hardware implementations of arithmetic operators with integrated fault detection (and/or fault tolerance) capabilities. The target arithmetic operators are: adders, subtracters, multipliers (and variants of multiplications by constants, square, FMA, MAC), division, square-root, approximations of the elementary functions. We study two approaches: residue codes and specific bit-level coding in some redundant number systems for fault detection/tolerance integration at the arithmetic operator/unit level. FPGA prototypes are under development.

**Secure Virtualization in Hardware**

In the paper [70] presented at SDTA, we deal with secure solutions that can help virtualization and communication which can be implemented on new hybrids (Core + FPGA) development platforms. On one side, these boards are featured with processors that do not have virtualization extensions but are powerfull enough to really support hypervisors and their guests. On the other side some virtualization solutions presently exist for ARM processors but they only refer to TrustZone for their (hardware) security. These hybrid boards can offer us more: we have read some recents and up-to-date specifications made by a consortium to help the implementation of hardware security. In this area, FPGA can help in securing virtualization. But we must notice that, for now, all has been made for Intel/AMD architectures and for a lone operating system. Even so, the whole propositions are too complex to be implemented on embedded systems. So, we will have to use some capabilities in hardware development and make software rearrangements to help us to design a functional solution.

## 6.3. Compilation and Synthesis for Reconfigurable Platform

### *6.3.1. Numerical Accuracy Analysis and Optimization*

**Participants:** Olivier Sentieys, Steven Derrien, Romuald Rocher, Pascal Scalart, Tomofumi Yuki, Aymen Chakhari, Gaël Deest.

The problem of accuracy evaluation is one of the most time consuming tasks during the fixed-point refinement process. Analytical techniques based on perturbation theory have been proposed in order to overcome the need for long fixed-point simulation. However, these techniques are not applicable in the presence of certain operations classified as un-smooth operations. In such circumstances, fixed-point simulation should be used. In [33], an algorithm detailing the hybrid technique which makes use of an analytical accuracy evaluation technique used to accelerate fixed-point simulation was proposed. This technique is applicable to signal processing systems with both feed-forward and feedback interconnect topology between its operations. The proposed algorithm makes use of the classification of operators as smooth or un-smooth and uses the analytical SNS model obtained by using our previously published analytical techniques to evaluate the impact of finite precision on smooth operators, while performing simulation of the un-smooth operators during fixed-point simulation. In other words, parts of the system are selectively simulated only when un-smooth errors occur and not otherwise. Thus, the effort for fixed-point simulation is greatly reduced. The acceleration obtained as a result of applications of the proposed technique is consistent with fixed-point simulation, while reducing the time taken for fixed-point simulation by several orders of magnitude. The preprocessing overhead consists

of deriving the single-noise-source model, and it is often small in comparison to the time required for fixed-point simulation. The advantage of using the proposed technique is that the user need not spend time on characterizing the nonlinearities associated with un-smooth operations. Several examples from general signal processing, communication, and image processing domains are considered for evaluation of the proposed hybrid technique. The acceleration obtained is quantified as an improvement factor. Very high improvement factors indicate that the hybrid simulation is several orders of magnitude faster than classical fixed-point simulation.

One of the limitation of analytical accuracy technique is that they are based on a Signal Flow Graph Representation of the system to be analyzed. This SFG model is currently built-out of a source program by flattening its whole control-flow (including full loop unrolling) which raises significant accuracy analysis issues. To overcome these limitations, we have proposed [41] to adapt state of the art accuracy analysis techniques to take advantage of compact polyhedral program representations. Combining the two approaches provide a more general and scalable framework which significantly extends the applicability of accuracy models, enabling the analysis of complex image processing kernels operating on multidimensional data-sets.

An analytical approach was studied to determine accuracy of systems including unsmooth operators. An unsmooth operator represents a function which is not derivable in all its definition interval (for example the sign operator). The classical model is no longer valid since these operators introduce errors that do not respect the Widrow assumption (their values are often higher than signal power). So an approach based on the distribution of the signal and the noise was proposed. We focused on recursive structures where an error influences future decision (such as Decision Feedback Equalizer). In that case, numerical analysis method (e.g., Newton Raphson algorithm) can be used. Moreover, an upper bound of the error probability can be analytically determined. We also studied the case of Turbo Coder and Decoder to determine data word-length ensuring sufficient system quality [17].

### *6.3.2. Reconfigurable Processor Extension Generation*

**Participants:** Christophe Wolinski, François Charot.

Most proposed techniques for automatic instruction sets extension usually dissociate pattern selection and instruction scheduling steps. The effects of the selection on the scheduling subsequently produced by the compiler must be predicted. This approach is suitable for specialized instructions having a one-cycle duration because the prediction will be correct in this case. However, for multi-cycle instructions, a selection that does not take scheduling into account is likely to privilege instructions which will be, *a posteriori*, less interesting than others in particular in the case where they can be executed in parallel with the processor core. The originality of our research work is to carry out specialized instructions selection and scheduling in a single optimization step. This complex problem is modeled and solved using constraint programming techniques. This approach allows the features of the extensible processor to be taken into account with a high degree of flexibility. Different architectures models can be envisioned. This can be an extensible processor tightly coupled to a hardware extension having a minimal number of internal registers used to store intermediate results, or a VLIW-oriented extension made up of several processing units working in parallel and controlled by a specialized instruction. These techniques have been implemented in the Gecos source-to-source framework.

Novel techniques addressing the interactions between code transformation (especially loops) and instruction set extension are under study. The idea is to automatically transform the original loop nests of a program (using the polyhedral model) to select specialized and vector instructions. These new instructions may use local memories located in the hardware extension and used to store intermediates data produced at a given loop iteration. Such transformations lead to patterns whose effect is to significantly reduce the pressure on the memory of the processor.

We also studied a way to identify custom instructions at the application domain level instead of addressing it on a per-application basis. Domain-specific instruction set extension aims at maximizing the usage of a custom instruction across a set of applications belonging to an application domain. The idea is to guarantee that each custom instruction has a high degree of utilization across many applications of a given domain,

while still delivering the required performance improvement. The instruction identification problem is here formulated as the maximum common subgraph problem and it is solved by transforming it into a maximum clique problem.

### 6.3.3. Optimization of Loop Kernels Using Software and Memory Information

**Participant:** Angeliki Kritikakou.

The compilers optimize the compilation sub-problems one after the other following an order which leads to less efficient solutions because the different sub-problems are independently optimized taking into account only a part of the information available in the algorithms and the architecture. In a paper accepted for publication in Computer Languages, Systems & Structures (COMLAN), Elsevier, we have presented an approach which applies loop transformations in order to increase the performance of loop kernels. The proposed approach focuses on reducing the L1, L2 data cache and main memory accesses and the addressing instructions. Our approach exploits the software information, such as the array subscript equations, and the memory architecture, such as the memory sizes. Then, it applies source-to-source transformations taking as input the C code of the loop kernels and producing a new C code which is compiled by the target compiler. We have applied our approach to five well-known loop kernels for both embedded processors and general purpose processors. From the obtained experimental results we observed speedup gains from 2 up to 18.

### 6.3.4. Design Tools for Reconfigurable Video Coding

**Participants:** Emmanuel Casseau, Yaset Oliva Venegas.

In the field of multimedia coding, standardization recommendations are always evolving. To reduce design time taking benefit of available SW and HW designs, Reconfigurable Video Coding (RVC) standard allows defining new codec algorithms. The application is represented by a network of interconnected components (so called actors) defined in a modular library and the behaviour of each actor is described in the specific RVC-CAL language. Dataflow programming, such as RVC applications, express explicit parallelism within an application. However general purpose processors cannot cope with both high performance and low power consumption requirements embedded systems have to face. We have investigated the mapping of RVC applications onto a dedicated multiprocessor platform. Actually, our goal is to propose an automated co-design flow based on the RVC framework. The designer provides the application description in the RVC-CAL language, after which the co-design flow automatically generates a network of processors that can be synthesized on FPGA platforms. Two kinds of platforms can be targeted. The first platform is made of processors based on a low complexity and configurable TTA processor (Very Long Instruction Word -style processor). The architecture model of the platform is composed of processors with their local memories, an interconnection network and shared memories. Both shared and local memories are used to limit the traditional memory bottleneck. Processors are connected together through the shared memories [72] [69] [36]. The second platform more specifically targets the Zynq platform from Xilinx. The processors are MicroBlaze processors. Their local memory is dedicated to instruction code only. A common shared memory is used for the data exchanges between the processors (to store the data that communicate between actors). At present time, the actor mapping is chosen at compile time but we expect dynamic mapping soon. The mapping will be computed at runtime on the ARM processor. The actor's code will be stored in the DDR memory so that it can be easily transferred to the MicroBlaze instruction cache depending on the actor mapping [55] [76] . This work is done in collaboration with IETR and has been implemented in the Orcc open-source compiler (Open RVC-CAL Compiler: http://orcc.sourceforge.net).

### 6.3.5. A Domain Specific Language for Rapid Prototyping of Software Radio Waveforms

**Participants:** Matthieu Gautier, Olivier Sentieys, Ganda-Stéphane Ouedraogo.

Software Defined Radio (SDR) is now becoming a ubiquitous concept to describe and implement Physical Layers (PHYs) of wireless systems. Moreover, even though the FPGA (Field Programmable Gate Array) technology is expected to play a key role in SDR, describing a PHY at the Register-Transfer-Level (RTL) requires tremendous efforts. We introduced a novel methodology to rapidly implement PHYs for FPGA-SDR

platforms. The work relies upon High-Level Synthesis tools and dataflow modeling to infer an efficient system-level control unit for the application. The proposed software-based over-layer partly handles the complexity of programming an FPGA and integrates reconfigurable features. It consists essentially of a Domain-Specific Language (DSL) [60] that handles the complexity of programming an FPGA and a DSL-Compiler [32] for automation purpose. IEEE 802.11a a and IEEE 802.15.4 transceivers have been designed and explored [45] via this new methodology in order to show the rapid prototyping feature.

## 6.4. Interaction between Algorithms and Architectures

### *6.4.1. Cooperative-cum-Constrained Maximum Likelihood Algorithm for UWB-based Localization in Wireless BANs*

**Participants:** Antoine Courtay, Matthieu Gautier, Gia Minh Hoang [Master's Student].

Wireless Body Area Network (BAN) is a mainstream technology for numerous application fields (medicine, security, sport science, etc.) and precise determination of wireless sensors' positions responses to the great needs in many applications. This study leverages Ultra Wide Band (UWB) radio which is an attractive technology to achieve the centimeter-level distance measurements. However, the aggregation of the distance information remains a challenge to achieve an accurate localization in wireless BAN. To this aim, we have proposed a novel Cooperative-cum-Constrained Maximum Likelihood (CCML) localization algorithm. This algorithmic study shows the improvement that could be achieved by combining UWB radio and dedicated algorithms. Future works is to integrate UWB technology in the second version of the Zyggie platform developed in CAIRN.

### *6.4.2. MIMO Systems and Cooperative Strategies for Low-Energy Wireless Networks*

**Participants:** Olivier Berder, Olivier Sentieys, Baptiste Vrigneau, Viet-Hoa Nguyen.

Since a couple of years, the CAIRNteam has reached a significant expertise in multi-antenna systems, especially in linear precoding. If this technique is traditionally used in a collocated way, it could also be used for wireless sensor networks (WSN) in a distributed manner. We presented a new approach, named distributed max-dmin precoding (DMP). This protocol is based on the deployment of a virtual 2x2 max-dmin precoding over one source, one forwarding relay, both equipped with one antenna and a destination involving two antennas. In this context, two kinds of relaying, amplify and forward or decode and forward protocols, were investigated. The performance evaluation in terms of Bit-Error-Rate (BER) and energy efficiency was compared with non cooperative techniques (SISO, SIMO) and the distributed space time block code (STBC) scheme. Our investigations showed that the DMP takes the advantage in terms of energy efficiency from medium transmission distances.

A receiver initiated cooperative medium access control (RIC-MAC) protocol was also proposed for cooperative communications to reduce the energy consumption of WSN. Considering a real WSN platform, the simulation results show that using the proposed RIC-MAC protocol in cooperative communications provides latency and energy gains as compared to multi-hop communications. Even if the energy gain is shown to be reduced when the network traffic load increases, our protocol still brings an energy gain about 22% at 1 packet/second. Finally, considering the impact of traffic load on energy consumption and latency, RIC-MAC is illustrated to be robust to traffic load variations in terms of latency [66].

### *6.4.3. Adaptive protocols for Wireless Sensor Networks*

**Participants:** Olivier Berder, Matthieu Gautier, Nhat-Quang Nhan [Master's Student], Van-Thiep Nguyen.

As tiny sensor nodes are equipped with limited battery, the optimization of the power consumption of these devices is extremely vital. In typical WSN platforms, the radio transceiver consumes major proportion of the energy. Major concerns are therefore to decrease the radio activity by designing efficient MAC protocols.

Energy consumption plays an important role in the design of Wireless Body Area Sensor Network (WBASN). Unfortunately, the performance of WBASNs decreases in high interference environments such as the Industrial, Scientific and Medical (ISM) band where wireless spectrums are getting crowded. In this study [59], an energy-efficient Medium Access Control (MAC) protocol named C-RICER (Cognitive-Receiver Initiated CyclEd Receiver) is specifically designed for WBASN to cognitively work in high interference environment. C-RICER protocol adapts both transmission power and channel frequency to reduce the interferences and thus, the energy consumption. The protocol is simulated with the OMNET++ simulator. Simulation results show that, depending on the interference level, C-RICER is able to outperform the traditional RICER protocol in terms of energy consumption, packet delay, and network throughput.

In recent years, many MAC protocols for Wireless Sensor Networks (WSNs) have been proposed and evaluated using Matlab simulator and/or network simulators (OMNeT++, NS2, etc.). However, most of them have a static behavior and few network simulations are available for adaptive protocols. Specially, in OMNeT++/MiXiM, there is few energy-efficient MAC protocol for WSNs (B-MAC and L-MAC) and no adaptive protocol. To this end, the TAD-MAC (Traffic Aware Dynamic MAC) protocol has been simulated in OMNeT++ with the MiXiM framework [57]. The simulation results have been used to compare with B-MAC and L-MAC protocol, showing the gain brought by TAD-MAC.

### 6.4.4. Energy Harvesting and Power Management

**Participants:** Olivier Berder, Olivier Sentieys, Arnaud Carer, Trong-Nhan Le.

To design autonomous Wireless Sensor Networks (WSNs) with a theoretical infinite lifetime, energy harvesting (EH) techniques have been recently considered as promising approaches. Ambient sources can provide everlasting additional energy for WSN nodes and exclude their dependence on battery. An efficient energy harvesting system which is compatible with various environmental sources such as light, heat or wind energy was proposed. Our platform takes advantage of double-level capacitors not only to prolong the system lifetime but also to enable robust booting from the exhausting energy of the system. Simulations and experiments showed that it can achieve booting time in order of seconds. Although capacitors have virtual recharge cycles, they suffer from higher leakage compared to rechargeable batteries. Increasing their size can decrease the system performance due to leakage energy. Therefore, an energy neutral design framework providing a methodology to determine the minimum size of the storage devices satisfying Energy Neutral Operation (ENO) and maximizing system Quality of Service (QoS) in EH nodes when using a given energy source was proposed. Experiments validating this framework were performed on a real WSN platform with both photovoltaic cells and thermal generators in an indoor environment [30].

A new PM for EH-WSNs scavenging energy from periodic sources, i.e., ambient energy is not available during the full harvesting cycle, was proposed. Not only respecting the ENO condition, our PM is able to balance the Quality of Service (QoS) during the whole cycle to provide regular data tracking, which is essential for WSN applications like monitoring. Simulations on OMNET++ show that our PM can improve the QoS during the absence of energy by a factor up to 84% compared to state-of-the-art PMs, while guaranteeing the same global QoS [54].

### 6.4.5. Multimedia Processing

**Participant:** Pascal Scalart.

Most noise reduction methods for multimedia signals are usually based on the application of a short-time Wiener filter (MMSE) that is generally expressed as a spectral gain depending on the local signal-to-noise ratio (SNR) on each frequency bin. To estimate such filter, several algorithms can be found in the literature but these conventional approaches lead to a biased estimator for the a priori signal-to-noise estimate. To reduce this bias, we have proposed in [26] a new strategy that relies on the introduction of a correction term in the computation of the Wiener filter depending on the current state of both the available a priori and a posteriori SNR estimates. The proposed solution leads to a bias-compensated a priori SNR estimate, and allows to finely estimating the target signal that is very close to the original noise-free reference. Such refinement procedure has been tested under various noisy environments and show the superiority of the proposed strategy compared to competitive algorithms.

Audio classification systems have recently gained interest for the design of various real-world multimedia services such as audio database indexing with musical genre classification, video indexing using the soundtrack or context awareness. A large majority of audio classification systems can be viewed as offline applications in the sense that there is no strong restriction about how the signal to be classified is accessed. In [44], we investigate the case where the classification task is performed in real-time in a low-latency classification framework. We proposed different methodologies for the use of feature integration that are based on three key aspects: the selection of the features which have to be temporally integrated, the choice of the integration techniques, i.e. how the temporal information is extracted, and the size of the integration window. The experiments carried out for the classification task show that these different methodologies have a significant impact on the global performance even with the low-latency constraints. In addition, we investigate the detection of howlings that arise in audio signals in [43]. To do so, the processing algorithm is based on a Support Vector Machine (SVM) model in the decision stage and on the combination of energy-based features and also a new feature related to the frequency stability of a howling component. The proposed method can be used in different situation since its provides good results with a very low false alarm rate for a wide range of experimental conditions.

### *6.4.6. Non-Intrusive Load Monitoring*

**Participants:** Olivier Sentieys, Baptiste Vrigneau, Xuan Chien Le.

Natural resource preservation has recently become a significant concern and has therefore motivated many research and development efforts for energy consumption management in buildings and homes. Efficiently reducing energy consumption at home, work or in a factory, could be afforded by mixing different technologies to not only reduce the energy consumed by consumers, but also to adapt (manage) the energy consumed to the energy that is produced. SMART 2020 outlined the opportunity to capture savings of both energy and Greenhouse Gas (GHG) emissions in 2020, through a range of actions developed by the Information and Communications Technologies (ICT) sector. Smart Grid, Smart Buildings, and Green ICT have the main impact on energy savings. At the energy production side, the electrical grid infrastructure is comprised of three elements: power generation, transmission, and distribution. Electrical power generation consists mainly of the power plants but also includes more and more renewable sources such as wind power or solar panels on energy farms or locally on top of buildings. The cost of energy storage is very high, and hence the current practice is to match energy consumption closely with energy generation, which is more and more fluctuating: challenges could be seen as being able to use energy when the wind blows or the sun shines, and also to avoid the strong power consumption peaks due to people's life. A typical example at home could be to automatically use the dryer when energy is available and therefore cheap, and is now well defined as Smart Grid technologies. At the energy consumption side, the main objective is of course to reduce energy consumption of the different subsystems. Interior lighting, office equipment, heating, cooling, and ventilation make up of more than 85% of the total electricity use and the reduction effort should therefore be concentrated on these systems. For energy management and reduction in homes or building a key enabler is the use of wireless sensor networks to monitor the environment (temperature, activity of people, power consumption of equipment, light, etc.) and to act on subsystems (decrease room temperature, stop or start an equipment, adjust cooling or ventilation, etc.). This is the emerging field of Smart Building Automation.

The objective of this work is strongly linked to the usage of these WSN nodes in the context of smart monitoring of energy consumption and environment (temperature, activity, light). We will propose new Indirect Power Monitoring techniques which enable to estimate energy consumed in a building or in a home without effectively measuring the power consumed. A typical AC smart meter is costly equipment and we therefore want to propose cheap and non-invasive sensor nodes. As an example, to estimate the power consumed by the TV, it is not necessary to measure precisely the current it consumed, but a simple sensor able to recognize that TV is on or off can do the same job with a far less complexity. Another example is the development and deployment of room occupancy and people activity sensors that can lead to significant reduction of the energy by regulating HVAC (Heating, Ventilation and Air-Conditioning) or by switching lights and office equipment. The wireless transmission is the main reason of consuming energy and the new algorithms will propose to make the sensors to cooperate inside a low-distance cluster (an office for example). The algorithms will decide the best strategy and the best information to send back in order to offer the best

trade-off between Performance/Complexity/Consumption. This work is closely links to power management techniques and energy harvesting (in-door light, heat, vibration). A power manager embedded in energy harvesting WSN nodes adapts the power consumption and computation loads according to the harvested energy to obtain a theoretically infinite lifetime. The main advantage of using energy harvesting (EH) in the context of building and home monitoring is to avoid battery replacement and therefore to reduce installation and maintenance costs of the system.

**CELTIQUE Project-Team**

# 5. New Results

## 5.1. Browser randomization against web tracking

**Participants:** Frédéric Besson, Thomas Jensen.

We have investigated different approaches for dynamically tracking information flows in order to improve web browser security. We have identified the problem of stateless web tracking (fingerprinting) and have proposed a novel approach to hybrid information flow monitoring by tracking the knowledge about secret variables using logical formulae. In a follow-up work we investigated how to enforce browser anonymity in the presence of finger-printing web trackers. One way to protect the users' privacy is to make them switch between different machine and browser configurations. We propose a formalisation of this privacy enforcement mechanism. We use information-theoretic channels to model the knowledge of the tracker and the fingerprinting program, and show how to synthesise a randomisation mechanism that defines the distribution of configurations for each user. This mechanism provides a strong guarantee of *privacy* (the probability of identifying the user is bounded by a given threshold) while maximising *usability* (the user switches to other configurations rarely). To find an optimal solution, we express the enforcement problem of randomisation by a linear program. We investigate and compare several approaches to randomisation and find that more efficient privacy enforcement would often provide lower usability. Finally, we relax the requirement of knowing the fingerprinting program in advance, by proposing a randomisation mechanism that guarantees privacy for an arbitrary program.

## 5.2. Static analysis of functional programs using tree automata and term rewriting

**Participants:** Thomas Genet, Barbara Kordy, Yann Salmon.

We develop a specific theory and the related tools for analyzing programs whose semantics is defined using term rewriting systems. The analysis principle is based on regular approximations of infinite sets of terms reachable by rewriting. The tools we develop use, so-called, Tree Automata Completion to compute a tree automaton recognizing a superset of all reachable terms. This over-approximation is then used to prove properties on the program by showing that some "bad" terms, encoding dangerous or problematic configurations, are not in the superset and thus not reachable. This is a specific form of, so-called, Regular Tree Model Checking. However, when dealing with infinite-state systems, Regular Tree Model Checking approaches may have some difficulties to represent infinite sets of data. We proposed Lattice Tree Automata, an extended version of tree automata to represent complex data domains and their related operations in an efficient manner. Moreover, we introduce a new completion-based algorithm for computing the possibly infinite set of reachable states in a finite amount of time. This algorithm is independent of the lattice making it possible to seamlessly plug abstract domains into a Regular Tree Model Checking algorithm. These results are part of Valérie Murat's PhD thesis [13]. Now, we aim at applying this technique to the static analysis of programming languages whose semantics is based on terms, like functional programming languages. We already shown that static analysis of first order functional programs can be automated using tree automata completion [28]. Now, one of the objective is to lift those results to the static analysis of higher-order functions. This was precisely the purpose of Yann Salmon's visit to Pr. Luke Ong. Barbara Kordy who joined Celtique in September 2014 is also going to work on this subject.

## 5.3. Certified JavaScript

**Participants:** Martin Bodin, Alan Schmitt.

We have completed our first milestone in the development of a certified JavaScript semantics. We have finished a first version of JSCert, a formalization of the current ECMA standard in the Coq proof assistant, and JSRef, a reference interpreter for JavaScript extracted from Coq to OCaml. We have also given a Coq proof that JSRef is correct with respect to JSCert and assessed JSRef using test262, the ECMA conformance test suite. Our methodology ensures that JSCert is a comparatively accurate formulation of the English standard. We have demonstrated that modern techniques of mechanized specification can handle the complexity of JavaScript. This result, obtained in the setting of a collaboration with Philippa Gardner and Sergio Maffeis of Imperial College, and Arthur Charguéraud of Inria Saclay, have been published in the conference Principles of Programming Languages [25].

## 5.4. SawjaCard: a static analysis tool for certifying Java Card applications

**Participants:** Frédéric Besson, Thomas Jensen, David Pichardie, Delphine Demange.

We have transfered to the FIME company a static analysis tool for certifying *Java Card* applications, according to security rules defined by the smart card industry. *Java Card* is a dialect of Java designed for programming multi-application smart cards and the tool, called *SawjaCard*, has been specialised for the particular *Java Card* programming patterns. The tool is built around a static analysis engine which uses a combination of numeric and heap analysis. It includes a model of the *Java Card* libraries and the *Java Card* firewall. The tool has been evaluated on a series of industrial applets and is shown to automate a substantial part of the validation process [21].

## 5.5. Semantics for C programs

**Participants:** Frédéric Besson, Sandrine Blazy, Pierre Wilke.

Real life C programs are often written using C dialects which, for the ISO C standard, have undefined behaviours. In particular, according to the ISO C standard, reading an uninitialised variable has an undefined behaviour and low-level pointer operations are implementation defined. We propose a formal semantics which gives a well-defined meaning to those behaviours for the C dialect of the CompCert compiler. Our semantics builds upon a novel memory model leveraging a notion of symbolic values. Symbolic values are used by the semantics to delay the evaluation of operations and are normalised lazily to genuine values when needed. We show that the most precise normalisation is computable and that a slightly relaxed normalisation can be efficiently implemented using an SMT solver. The semantics is executable and our experiments show that the enhancements of our semantics are mandatory to give a meaning to low-levels idioms such as those found in the allocation functions of a C standard library [21].

## 5.6. Fast inference of polynomial invariants

**Participants:** David Cachera, Thomas Jensen.

We have developed our static analysis techniques for computing polynomial invariants for imperative programs. The analysis is derived from an abstract interpretation of a backwards semantics, and computes preconditions for equalities of the form g = 0 to hold at the end of execution. A distinguishing feature of the technique is that it computes polynomial loop invariants without resorting to Gro¨bner base computations. The analysis uses remainder computations over parameterized polynomials in order to handle conditionals and loops efficiently. The algorithm can analyze and find a large majority of loop invariants reported previously in the literature, and executes significantly faster than implementations using Gro¨bner bases [15].

## 5.7. Quantitative analysis of security

**Participant:** Barbara Kordy.

Graphical models for security is a young but rapidly growing research field. Security models based on graphs combine intuitive, visual representation with rigorous, mathematical foundations. In [30] we address the growing need of performing meaningful probabilistic analysis of security using graphical models. We propose a framework that integrates the modeling technique of attack–defense trees with probabilistic information expressed in terms of Bayesian networks. This allows us to perform probabilistic evaluation of attack–defense scenarios involving dependent actions. To improve the efficiency of our computations, we make use of inference algorithms from Bayesian networks and encoding techniques from constraint reasoning. We discuss the algebraic theory underlying our framework and point out several generalizations which are possible thanks to the use of semiring theory

## 5.8. Formal Verification of an SSA-Based Middle-End for CompCert

**Participants:** Delphine Demange, David Pichardie.

CompCert is a formally verified compiler that generates compact and efficient code for a large subset of the C language. However, CompCert foregoes using SSA, an intermediate representation employed by many compilers that enables writing simpler, faster optimizers. In fact, it has remained an open problem to verify formally an SSA-based compiler. We report in [14] on a formally verified, SSA-based middle-end for CompCert. In addition to providing a formally verified SSA-based middle-end, we address two problems raised by Leroy in 2009: giving an intuitive formal semantics to SSA, and leveraging its global properties to reason locally about program optimizations. Joint work with Gilles Barthe.

## 5.9. A verified information-flow architecture

**Participants:** Delphine Demange, David Pichardie.

SAFE is a clean-slate design for a highly secure computer system, with pervasive mechanisms for tracking and limiting information flows. At the lowest level, the SAFE hardware supports fine-grained programmable tags, with efficient and flexible propagation and combination of tags as instructions are executed. The operating system virtualizes these generic facilities to present an information-flow abstract machine that allows user programs to label sensitive data with rich confidentiality policies. We present a formal, machine-checked model of the key hardware and software mechanisms used to control information flow in SAFE and an end-to-end proof of noninterference for this model in the Coq proof assistant [17]. This work has been obtained in collaboration with colleagues from University of Pennsylvania, Portland State University, and Harvard University, as part of the CRASH-SAFE project, funded by DARPA.

## 5.10. Formal Verification of Static Analysis

**Participants:** Sandrine Blazy, Vincent Laporte, David Pichardie.

Static analysis of binary code is challenging for several reasons. In particular, standard static analysis techniques operate over control flow graphs, which are not available when dealing with self-modifying programs which can modify their own code at runtime. We formalized in the Coq proof assistant some key abstract interpretation techniques that automatically extract memory safety properties and control flow graphs from binary code [22], and operate over a small subset of the x86 assembly. Our analyzer is formally proved correct and has been run on several self-modifying challenges, provided by Cai et al. in their PLDI 2007 paper.

**CIDRE Project-Team**

# 6. New Results

## 6.1. Highlights of the Year

The supervision of distributed system relies heavily on correlation mechanisms that are responsible for collecting alerts coming from sensors and detecting complex scenarios in the flow of alerts. The problem is that it requires to write complex correlation rules. The work we have performed proposes a technique to generate semi-automatically such correlation rules. It describes a process that uses an attack tree and a representation of the system as inputs, and generate a correlation tree that can be translated in an alert correlation description language. This work received the best paper award of SAR-SSI 2014 [50].

One approach to protect the privacy of users in personalized recommendation systems is to publish a sanitized version of the profile of the user by relying a non-interactive mechanism compliant with the concept of differential privacy. In a joint work with Raghavendran Balu and Teddy Furon (LinkMedia Inria team), we have consider two existing schemes offering a differentially private representation of profiles: BLIP (BLoom-and-flIP) and JLT (Johnson-Lindenstrauss Transform). For assessing their security levels, we play the role of an adversary aiming at reconstructing a user profile. To realize this, we design two inference attacks named single and joint decoding. The first inference attack tests the presence of a single item in the profile, and is iterated independently for each possible item of the item set. In contrast, the second inference attack aims at deciding whether a particular subset of items is likely to be in the user profile. This attack is tested on all the possible subsets of items. Our contributions are a theoretical analysis and practical implementations of both attacks tested on datasets composed of real user profiles revealing that joint decoding is the most powerful attack. This also gives useful insights on the setting the differential privacy parameter $\epsilon$. This work has received the best student paper award at the conference ESORICS 2014.
BEST PAPERS AWARDS :
[27] **European Symposium on Research in Computer Security**. R. BALU, T. FURON, S. GAMBS.

## 6.2. Intrusion Detection

### *6.2.1. Intrusion detection based on an analysis of information flow control*

In 2014, Laurent Georget has started his PhD thesis in the team, working on a subject related to the analysis of information flow control at the kernel level. The goal of his PhD thesis is to propose a formal semantics of the system calls for a real operating systems (namely Linux). This semantics will provide insights about these system calls in terms of information flow. This work will help us to test in a more systematic and efficient way, our reference implementation of a information monitor at the kernel level (Blare).

Blare allows monitoring information flow and identifies the flows that do not conform to a security policy that has been previously defined. Please notice that any explicit flows between OS objects (sockets, files, etc.) are monitored and that in consequence hidden channel attacks cannot be detected by this approach.

We have already developed a dedicated test framework for this software. However, each test written by the developer must be accompanied with the possible results in terms of information flows. The framework simply compares the effective result with the set of expected results. A test passes when the effective result belongs to the set of expected results, and fails otherwise. However, this strategy has turned to be less intuitive than expected. Some system calls must be tested by using several processes operating concurrently. In these cases, the scheduling of processes can produce many different scenarios that will translate quite differently in terms of information flows. To be more confident in our implementation, we really need a stronger and more formal path. The PhD thesis of Laurent Georget is trying to bridge the gap between Blare implementation and the interpretation of the results obtained by running the information flow monitor.

### 6.2.2. Malware characterization through information flow monitoring

Monitoring information flows consists in observing how pieces of information are disseminated in a given environment. At system level, it consists in intercepting actions performed by an application to deduce how the application disseminates information within the entire operating system. We have propose a new approach to classify and later detect applications infected by malware based on the way they disseminate their own data within an operating system. For this purpose, we first introduce a data-structure named System Flow Graph [thèse Rado to ref.] that offers a compact representation of how pieces of data flow inside a system. A system flow graph describes the external behavior of an application during one execution. Its construction requires no knowledge about the inner working of the application. The graph is built using Blare as an information-flow monitor and more precisely its produced log. We have presented in [25] how these graphs reveal helpful to understand malware behavior and thus why it can help an expert to give a diagnosis in case of intrusion.

### 6.2.3. Terminating-insensitive non-interference verification based on information flow control

In 2010-2011, we started an informal collaboration with colleagues from CEA LIST laboratory. This collaboration has turned into a reality by the funding of a PhD student (Mounir Assaf). This PhD thesis is about the verification of security properties of programs written in an imperative language with pointer aliasing (a subset of C language) by techniques borrowed from the domain of static analysis. One of the property of interest for the security field is called terminating-insensitive non-interference. Briefly speaking, when verified by a program, this property ensures that the content of any secret variable can not leak into public ones (for any terminating execution). However, this property is too strict in the sense that a large number of programs although perfectly secure are rejected by classical analyzers. Finally in 2014, Mounir Assaf enhanced his previous work on static analysis by introducing a method permitting to quantify information leakage in a C program. This approach requires a theoretical definition of the quantification of information flow leakage and is very promising.

### 6.2.4. Visualization of security events

The first part of this year was dedicated to tune a working prototype of ELVIS [38] in order to perform field trials with our partner DGA-MI. The prototype was largely well accepted. We were invited by the DGA-MI to present a poster in the Forum DGA Innovation 2014. We will also present ELVIS during the FIC 2014 in Lille on the Pôle Cyber-Défense area.

However, ELVIS also exhibited some limitations of our approach in the way multiple datasets are handled together. We therefore went for a new cycle of research whose objective is to enhance ELVIS in two ways: first to handle multiple datasets at the same time, and second to improve interactions so as to better fit with the processes in forensics. The results of our research lead to CORGI (Combination, Organization and Reconstruction through Graphical Interactions) [39] which was presented at VizSec 2014 (part of Vis 2014). CORGI improves ELVIS by introducing the concepts of *values of interest* that consist in interesting values found by an analyst and that can be used later to search and filter in the other datasets. They are an intuitive and efficient way to link various datasets while the analyst performs its tasks. An early prototype has been developed.

### 6.2.5. Control flow integrity

In [40] we have studied physical attacks that could disturb the normal execution of an embedded program of a smartcard. Such attacks can be performed using laser beams, electromagnetic glitches and can corrupt the flow of information or change the control flow of the program. We have studied the particular case of the control flow and we have developed software countermeasures that increase the robustness of the control flow. These countermeasures do not require any additional software or hardware external components which is useful for devices like smartcards whose architecture cannot be modified. The developed countermeasures have been validated with the help of the VIS model checker in order to verify that they do not disturb the original execution of the code.

### 6.2.6. Alert correlation in distributed systems

In large systems, multiple (host and network) Intrusion Detection Systems (IDS) and many sensors are usually deployed. They continuously and independently generate notifications (event's observations, warnings and alerts). To cope with this amount of collected data, alert correlation systems have to be designed. An alert correlation system aims at exploiting the known relationships between some elements that appear in the flow of low level notifications to generate high semantic meta-alerts. The main goal is to reduce the number of alerts returned to the security administrator and to allow a higher level analysis of the situation. However, producing correlation rules is a highly difficult operation, as it requires both the knowledge of an attacker, and the knowledge of the functionalities of all IDSes involved in the detection process. In [50], [47], [36], we focus on the transformation process that allows to translate the description of a complex attack scenario into correlation rules. We show that, once a human expert has provided an action tree derived from an attack tree, a fully automated transformation process can generate exhaustive correlation rules that would be tedious and error prone to enumerate by hand. The transformation relies on a detailed description of various aspects of the real execution environment (topology of the system, deployed services, etc.). Consequently, the generated correlation rules are tightly linked to the characteristics of the monitored information system. The proposed transformation process has been implemented in a prototype that generates correlation rules expressed in an attack description language called Adele.

In the context of the PhD of Mouna Hkimi, we propose a approach to detect intrusions that affect the behavior of distributed applications. To determine whether an observed behavior is normal or not (occurrence of an attack), we rely on a model of normal behavior. This model has been built during an initial training phase. During this preliminary phase, the application is executed several times in a safe environment. The gathered traces (sequences of actions) are used to generate an automaton that characterizes all these acceptable behaviors. To reduce the size of the automaton and to be able to accept more general behaviors that are close to the observed traces, the automaton is transformed. These transformations may lead to introduce unacceptable behaviors. Our current work aims at identifying the possible errors tolerated by the compacted automaton.

## 6.3. Privacy

### 6.3.1. Privacy in location-based services

With the advent of GPS-equipped devices, a massive amount of location data is being collected, raising the issue of the privacy risks incurred by the individuals whose movements are recorded. In [17], we focus on a specific inference attack called the de-anonymization attack, by which an adversary tries to infer the identity of a particular individual behind a set of mobility traces. More specifically, we propose an implementation of this attack based on a mobility model called Mobility Markov Chain (MMC). A MMC is built out from the mobility traces observed during the training phase and is used to perform the attack during the testing phase. We design several distance metrics quantifying the closeness between two MMCs and combine these distances to build de-anonymizers that can re-identify users in an anonymized geolocated dataset. Experiments conducted on real datasets demonstrate that the attack is both accurate and resilient to sanitization mechanisms such as downsampling.

One example of a location-based services is dynamic carpooling (also known as instant or ad-hoc ridesharing), which is a service that arranges one-time shared rides on very short notice. This type of carpooling generally makes use of three recent technological advances: (i) navigation devices to determine a route and arrange the shared ride; (ii) smartphones for a traveller to request a ride from wherever she happens to be; and (iii) social networks to establish trust between drivers and passengers. However, the mobiquitous environment in which dynamic carpooling is expected to operate raises several privacy issues. Among all the personal identifiable information, learning the location of an individual is one of the greatest threats against her privacy. For instance, the spatio-temporal data of an individual can be used to infer the location of her home and workplace, to trace her movements and habits, to learn information about her centre of interests or even to detect a change from her usual behavior. Therefore, preserving location privacy is a major issue to be able to leverage the possibilities offered by dynamic carpooling. In a joint work with researchers from LAAS-CNRS

[16], we have propose to follow the privacy-by-design approach by integrating the privacy aspect in the design of dynamic carpooling, henceforth increasing its public (and political) acceptability and trust.

A secure location-based service requires that a mobile user certifies his position before gaining access to a resource. Currently, most of the existing solutions addressing this issue assume a trusted third party that can vouch for the position claimed by a user. However, as computation and communication capacities become ubiquitous with the large scale adoption of smartphones by individuals, these resources can be leverage on to solve this issue in a collaborative and private manner. More precisely together with researchers from LAAS-CNRS, we introduce PROPS, for Privacy-Preserving lOcation Proof System, which allows users to generate proofs of location in a private and distributed way using neighboring nodes as witnesses [35]. PROPS provides security properties such as unforgeability and non-transferability of the proofs, as well as resistance to classical localization attacks.

One of the fundamental building block to construct a location proof system such as PROPS is a distance-bounding protocol. More precisely, in distance-bounding authentication protocols a verifier assesses that a prover is (1) legitimate and (2) in the verifier's proximity. Proximity checking is done by running time-critical exchanges between both parties. This enables the verifier to detect relay attacks (also called mafia fraud). While most distance-bounding protocols offer resistance to mafia, distance, and impersonation attacks, only few protect the privacy of the authenticating prover. One exception is the protocol due to Hermans, Peeters, and Onete, which offers prover untraceability with respect to a Man-in-the-Middle adversary. However in this protocol as well as in all other distance-bounding protocols, any legitimate verifier can identify, and thus track, the prover. In order to counter the threats of possible corruption or data leakage from verifiers, together with Jean-Marc Robert (ETS, Montréal) we propose a distance-bounding protocol providing strong prover privacy with respect to the verifier and deniability with respect to a centralized back-end server managing prover creation and revocation [33]. In particular, we first formalize the notion of prover anonymity, which guarantees that even verifiers cannot trace provers, and deniability, which allows provers to deny that they were authenticated by a verifier. Finally, we prove that our protocol achieves these strong guarantees.

A particular class of relay attacks against distance-bounding protocols is called terrorist fraud in which a distant malicious prover colludes with an attacker located in a verier's proximity when authenticating. Existing distance-bounding protocols resisting such attacks are designed to be lightweight and thus symmetric, relying on a secret shared by the prover and the verifier. Recently, several asymmetric distance-bounding protocols were proposed by Gambs, Onete and Robert as well as by Hermans, Peter and Onete, but they fail to thwart terrorist fraud. One earlier asymmetric protocol aiming to be terrorist-fraud resistant is the DBPK-Log protocol due to Bussard and Bagga, which was unfortunately recently proven to achieve neither distance- nor terrorist-fraud resistance. In this work, we build on some ideas of the DBPK-Log scheme and propose a novel distance-bounding protocol resistant to terrorist fraud that does not require the pre-existence of a shared secret between the prover and the verifier [32]. Our construction, denoted as VSSDB (for Verifiable Secret Sharing and Distance-Bounding Protocol) relies on a veriable secret sharing scheme and on the concept of modes, which we introduce as a novel element to complement fast-round challenges in order to improve security. We prove that VSSDB achieves terrorist-fraud resistance in a relaxed security model called KeyTF-security, which we also present in this paper.

### *6.3.2. Equity in privacy-enhanced social networks*

In [46], we have examined a novel issue in the field of policy conflict resolution, and applied it to privacy policy management in distributed social networking systems. We accepted as a starting point that in a privacy-enhanced social network, when a user publishes a document (e.g., a picture), any user referenced in this document (e.g., people tagged in pictures) should be entitled to issue a privacy policy over this document. In this case, when a given user tries to access a given document, multiple users may issue multiple access control decisions (or rulings), possibly resulting in a normative conflict. Quite a number of strategies are available for the resolution of such conflicts, the most common one being the "deny strategy", allowing any ruling denying access to the resource to take precedence over others. This is usually considered a "secure" way of dealing with access control. However, with this strategy as with many others, it is possible for a user to design her policy in a way that systematically prevents other users from interacting in a normal way, while allowing herself to

potentially benefit from other people's more flexible policies. This may leads to unfair situations, in which some users take advantage of the systems while others' experience is damaged. This is particularly an issue in social networking applications, in which information sharing is a core feature and access restrictions, while necessary to protect intimacy, can sometimes be considered aggressive.

To address this particular trade-off between privacy and usability, we have introduced the notion of equity in such scenarios, a situation being equitable when all involved users have seen their policy enforced or violated in the same proportion over past interactions. We have designed a conflict resolution algorithm aimed at improving this equity in our social networking scenario, and evaluated its impact by measuring Gini coefficients (an indicator commonly used by economists to measure the distribution of wealth in a population) over the distribution of enforcement proportions in the population of users. With respect to this criterion, it actually proved more efficient than other strategies. Following these positive results, we have recently taken steps towards a formalization and generalization of this intuitive concept of equity and the design of systematic tools to evaluate and compare the impact of any conflict resolution strategy over various possible flavors of the notion.

### *6.3.3. Private mobile services*

The development of NFC-enabled smartphones has paved the way to new applications such as mobile payment (m-payment) and mobile ticketing (m-ticketing). However, often the privacy of users of such services is either not taken into account or based on simple pseudonyms, which does not offer strong privacy properties such as the unlinkability of transactions and minimal information leakage. In [48], [15], we introduce a lightweight privacy-preserving contactless transport service that uses the SIM card as a secure element. Our implementation of this service uses a group signature protocol in which costly cryptographic operations are delegated to the mobile phone. We have also conducted an interdisciplinary study with researchers from social sciences to analyze the media coverage in the modern public space on the topic of privacy with respect to mobile technologies [29]. Despite the difficulties highlighted by these studies, we argue that research efforts should support the emergence of mobile services that respect users' privacy as well as the development of a digital culture of privacy.

### *6.3.4. Architectures for privacy*

In the current architecture of the Internet, there is a strong asymmetry in terms of power between the entities that gather and process personal data (e.g., major Internet companies, telecom operators, cloud providers, ...) and the individuals from which this personal data is issued. In particular, individuals have no choice but to blindly trust that these entities will respect their privacy and protect their personal data. In a position paper [34] in a collaboration with researchers from the Université de Montréal and Aarhus University, we propose an utopian crypto-democracy model based on existing scientific achievements from the field of cryptography. More precisely, our main objective is to show that cryptographic primitives, including in particular secure multiparty computation, offer a practical solution to protect privacy while minimizing the trust assumptions. In the crypto-democracy envisioned, individuals do not have to trust a single physical entity with their personal data but rather their data is distributed among several institutions. Together these institutions form a virtual entity called the Trustworthy that is responsible for the storage of this data but which can also compute on it (provided first that all the institutions agree on this). Finally, we also propose a realistic proof-of-concept of the Trustworthy, in which the roles of institutions are played by universities. This proof-of-concept would have an important impact in demonstrating the possibilities offered by the crypto-democracy paradigm.

Active fingerprinting schemes were originally invented to deter malicious users from illegally releasing an item, such as a movie or an image. To achieve this, each time an item is released, a different fingerprint is embedded in it. If the fingerprint is created from an anti-collusion code, the fingerprinting scheme can trace colluding buyers who forge fake copies of the item using their own legitimate copies. Charpentier, Fontaine, Furon and Cox were the first to propose an asymmetric fingerprinting scheme based on Tardos codes, the most efficient anti-collusion codes known to this day. However, their work focuses on security but does not preserve the privacy of buyers. To address this issue, we introduce the first privacy-preserving asymmetric fingerprinting protocol based on Tardos codes [30]. This protocol is optimal with respect traitor tracing. We

also formally define the properties of correctness, anti-framing, traitor tracing, as well as buyer- and item-unlinkability. Finally, we prove that our protocol achieves these properties and give exact bounds for each of them.

### 6.3.5. Privacy and web services

We have proposed [61] a new model of security policy based for a first part on our previous works in information flow policy and for a second part on a model of Myers and Liskov. This new model of information flow serves web services security and allows a user to precisely define where its own sensitive pieces of data are allowed to flow through the definition of an information flow policy. A novel feature of such policy is that they can be dynamically updated, which is fundamental in the context of web services that allow the dynamic discovery of services. We have also presented an implementation of this model in a web services orchestration in BPEL (Business Process Execution Language).

### 6.3.6. Privacy-preserving ad-hoc routing

Last year, we have proposed NoName, a privacy-preserving ad-hoc routing protocol. Based on trapdoor, virtual switching and partially disjoint multipaths using Bloom filter, NoName ensures the anonymity of the source, of the destination and of intermediate nodes. It also ensures unlinkability between source and message and between destination and message. Since then, we have demonstrated that colluding attackers analyzing Bloom filters can locate the origin node of routes requests messages. Thus, Noname, like ARMR, another privacy-preserving ad-hoc routing protocol using also Bloom filter, do not prevent the localization of the source. We have developed a cryptographic primitive called fuzzy cryptographic Bloom filter that offers the same functions as Bloom filters (in our case, preventing routing loops) while preventing localization of the source of route request messages.

## 6.4. Trust

Digital reputation mechanisms have indeed emerged as a promising approach to cope with the specificities of large scale and dynamic systems. Similarly to real world reputation, a digital reputation mechanism expresses a collective opinion about a target user based on aggregated feedback about his past behavior. The resulting reputation score is usually a mathematical object (*e.g.* a number or a percentage). It is used to help entities in deciding whether an interaction with a target user should be considered. Digital reputation mechanisms are thus a powerful tool to incite users to behave trustworthily. Indeed, a user who behaves correctly improves his reputation score, encouraging more users to interact with him. In contrast, misbehaving users have lower reputation scores, which makes it harder for them to interact with other users. To be useful, a reputation mechanism must itself be accurate against adversarial behaviors. Indeed, a user may attack the mechanism to increase his own reputation score or to reduce the reputation of a competitor. A user may also free-ride the mechanism and estimate the reputation of other users without providing his own feedback. From what has been said, it should be clear that reputation is beneficial in order to reduce the potential risk of communicating with almost or completely unknown entities. Unfortunately, the user privacy may easily be jeopardized by reputation mechanisms, which is clearly a strong argument to compromise the use of such a mechanism. Indeed, by collecting and aggregating user feedback, or by simply interacting with someone, reputation systems can be easily manipulated in order to deduce user profiles. Thus preserving user privacy while computing robust reputation is a real and important issue that we address in our work [51]. Specifically, our proposal aims at enhancing signatures of reputation mechanism proposed by Bethencourt and his colleagues in 2010 by handling negative votes. Taking into account negative votes implies major modifications with respect to the implementation of the mechanism. Specifically, in the mechanism of Bethencourt and co-authors, service providers locally store votes cast at the end of their interaction with their clients, and compute their reputation score by aggregating the received votes. In particular, they can keep only a subset of them, which clearly makes negative votes useless. We propose to improve upon this solution by guaranteeing that negative votes are taken into account. This is achieved by making both reputation scores and votes of service providers publicly available in order to prevent anyone from modifying or hiding them. Our proposition accomplishes this without jeopardizing the privacy of clients.

## 6.5. Other topics related to security and distributed computing

### 6.5.1. Network monitoring and fault detection

Monitoring a system consists in collecting and analyzing relevant information provided by the monitored devices, so as to be continuously aware of the system state (situational awareness). However, the ever growing complexity and scale of systems makes both real time monitoring and fault detection a quite tedious task. Thus the usually adopted option is to focus solely on a subset of information states, so as to provide coarse-grained indicators. As a consequence, detecting isolated failures or anomalies is a quite challenging issue. We propose in [23], [42] to address this issue by pushing the monitoring task at the edge of the network. We present a peer-to-peer based architecture, which enables nodes to adaptively and efficiently self-organize according to their "health" indicators. By exploiting both temporal and spatial correlations that exist between a device and its vicinity, our approach guarantees that only isolated anomalies (an anomaly is isolated if it impacts solely a monitored device) are reported on the fly to the network operator. We show that the end-to-end detection process, *i.e.*, from the local detection to the management operator reporting, requires a logarithmic number of messages in the size of the network.

### 6.5.2. Secure data deduplication scheme

Data grows at the impressive rate of 50% per year, and 75% of the digital world is a copy [0]. Although keeping multiple copies of data is necessary to guarantee their availability and long term durability, in many situations the amount of data redundancy is immoderate. By keeping a single copy of repeated data, data deduplication is considered as one of the most promising solutions to reduce the storage costs, and improve users experience by saving network bandwidth and reducing backup time. However, this solution must now solve many security issues to be completely satisfying. In this paper we target the attacks from malicious clients that are based on the manipulation of data identifiers and those based on backup time and network traffic observation. In [43], we have presented a deduplication scheme mixing an intra-and an inter-user deduplication in order to build a storage system that is secure against the aforementioned type of attacks by controlling the correspondence between files and their identifiers, and making the inter-user deduplication unnoticeable to clients using deduplication proxies. Our method provides global storage space savings, per-client bandwidth network savings between clients and deduplication proxies, and global network bandwidth savings between deduplication proxies and the storage server. The evaluation of our solution compared to a classic system shows that the overhead introduced by our scheme is mostly due to data encryption which is necessary to ensure confidentiality. This work relies on Mistore [44], [45], a distributed storage system aiming at guaranteeing data availability, durability, low access latency by leveraging the Digital Subscriber Line infrastructure of an ISP. Mistore uses the available storage resources of a large number of home gateways and points of presence for content storage and caching facilities reducing the role of the data center to a load balancer. Mistore also targets data consistency by providing multiple types of consistency criteria on content and a versioning system allowing users to get access to any prior versions of their contents.

### 6.5.3. Metrics estimation on very large data streams

In [12], we consider the setting of large scale distributed systems, in which each node needs to quickly process a huge amount of data received in the form of a stream that may have been tampered with by an adversary (*i.e.*, data items ordering can be manipulated by an oblivious adversary). In this situation, a fundamental problem is how to detect and quantify the amount of work performed by the adversary. To address this issue, we propose AnKLe (for Attack-tolerant eNhanced Kullback- Leibler divergence Estimator), a novel algorithm for estimating the KL divergence of an observed stream compared to the expected one. AnKLe combines sampling techniques and information-theoretic methods. It is very efficient, both in terms of space and time complexities, and requires only a single pass over the data stream. Experimental results show that the estimation provided by AnKLe remains accurate even for different adversarial settings for which the quality of other methods dramatically decreases. Considering $n$ as the number of distinct data items in a stream, we show that AnKLe is an $(\varepsilon, \delta)$-approximation algorithm with a space complexity sublinear in the size of the domain value from which data items are drawn and the maximal stream length.

[0] The digital universe decade. Are you ready? John Gantz and David Reinsel, IDC information, may 2010.

We go a step further by proposing in [22] a metric, called codeviation, that allows to evaluate the correlation between distributed streams. This metric is inspired from classical metric in statistics and probability theory, and as such allows us to understand how observed quantities change together, and in which proportion. We then propose to estimate the codeviation in the data stream model. In this model, functions are estimated on a huge sequence of data items, in an online fashion, and with a very small amount of memory with respect to both the size of the input stream and the values domain from which data items are drawn. We give upper and lower bounds on the quality of the codeviation, and provide both local and distributed algorithms that additively approximates the codeviation among $n$ data streams by using a sublinear number of bits of space in the size of the domain value from which data items are drawn and the maximal stream length. To the best of our knowledge, such a metric has never been proposed so far.

### 6.5.4. Robustness analysis of large scale distributed systems

In the continuation of [59] which proposed an in-depth study of the dynamicity and robustness properties of large-scale distributed systems, we analyze in [13], the behavior of a stochastic system composed of several identically distributed, but non independent, discrete-time absorbing Markov chains competing at each instant for a transition. The competition consists in determining at each instant, using a given probability distribution, the only Markov chain allowed to make a transition. We analyze the first time at which one of the Markov chains reaches its absorbing state. When the number of Markov chains goes to infinity, we analyze the asymptotic behavior of the system for an arbitrary probability mass function governing the competition. We give conditions for the existence of the asymptotic distribution and we show how these results apply to cluster-based distributed systems when the competition between the Markov chains is handled by using a geometric distribution.

### 6.5.5. Detection of distributed denial-of-service attacks

A Denial-of-Service (DoS) attack tries to progressively take down an Internet resource by flooding this resource with more requests than it is capable to handle. A Distributed Denial-of-Service (DDoS) attack is a DoS attack triggered by thousands of machines that have been infected by a malicious software, with as immediate consequence the total shut down of targeted web resources (*e.g.*, e-commerce websites). A solution to detect and to mitigate DDoS attacks it to monitor network traffic at routers and to look for highly frequent signatures that might suggest ongoing attacks. A recent strategy followed by the attackers is to hide their massive flow of requests over a multitude of routes, so that locally, these flows do not appear as frequent, while globally they represent a significant portion of the network traffic. The term “iceberg” has been recently introduced to describe such an attack as only a very small part of the iceberg can be observed from each single router. The approach adopted to defend against such new attacks is to rely on multiple routers that locally monitor their network traffic, and upon detection of potential icebergs, inform a monitoring server that aggregates all the monitored information to accurately detect icebergs. To prevent the server from being overloaded by all the monitored information, routers continuously keep track of the $c$ (among $n$) most recent high flows (modeled as items) prior to sending them to the server, and throw away all the items that appear with a small probability $p_i$, and such that the sum of these small probabilities is modeled by probability $p_0$. Parameter $c$ is dimensioned so that the frequency at which all the routers send their $c$ last frequent items is low enough to enable the server to aggregate all of them and to trigger a DDoS alarm when needed. This amounts to compute the time needed to collect $c$ distinct items among $n$ frequent ones. A thorough analysis of the time needed to collect $c$ distinct items appears in [53].

### 6.5.6. Randomized message-passing test-and-set

In [56], we have presented a solution to the well-known Test&Set operation in an asynchronous system prone to process crashes. Test&Set is a synchronization operation that, when invoked by a set of processes, returns yes to a unique process and returns no to all the others. Recently many advances in implementing Test&Set objects have been achieved, however all of them target the shared memory model. In this paper we propose an implementation of a Test&Set object in the message passing model. This implementation can be invoked by any number $p < n$ of processes in which $n$ is the total number of processes in the system. It has an expected individual step complexity in $O(\log p)$ against an oblivious adversary, and an expected

individual message complexity in $O(n)$. The proposed Test&Set object is built atop a new basic building block, called selector, that allows to select a winning group among two groups of processes. We propose a message-passing implementation of the selector whose step complexity is constant. We are not aware of any other implementation of the Test&Set operation in the message passing model.

### *6.5.7. Agreement problems in unreliable systems*

In [18], we consider the problem of approximate consensus in mobile ad-hoc networks in the presence of Byzantine nodes. Each node begins to participate by providing a real number called its initial value. Eventually all correct nodes must obtain final values that are different from each other within a maximum value previously defined (convergence property) and must be in the range of initial values proposed by the correct nodes (validity property). Due to nodes' mobility, the topology is dynamic and unpredictable. We propose an approximate Byzantine consensus protocol which is based on the linear iteration method. Each node repeatedly executes rounds. During a round, a node moves to a new location, broadcasts its current value, gathers values from its neighbors, and possibly updates its value. In our protocol, nodes are allowed to collect information during several consecutive rounds: thus moving gives them the opportunity to gather progressively enough values. An integer parameter $R_c$ is used to define the maximal number of rounds during which values can be gathered and stored while waiting to be used. A novel sufficient and necessary condition guarantees the final convergence of the consensus protocol. At each stage of the computation, a single correct node is concerned by the requirement expressed by this new condition (the condition is not universal as it is the case in all previous related works). Moreover the condition considers both the topology and the values proposed by correct nodes. If less than one third of the nodes are faulty, the condition can be satisfied. We are working on mobility scenarios (random trajectories, predefined trajectories, meeting points) to assert that the condition can be satisfied for reasonable values of $R_c$. In [41], we extend the above protocol to solve the problem of clock synchronization in mobile ad-hoc networks.

In [20], we investigate the use of agreement protocols to develop transactional mobile agents. Mobile devices are now equipped with multiple sensors and networking capabilities. They can gather information about their surrounding environment and interact both with nearby nodes, using a dynamic and self-configurable ad-hoc network, and with distant nodes via the Internet. While the concept of mobile agent is appropriate to explore the ad-hoc network and autonomously discover service providers, it is not suitable for the implementation of strong distributed synchronization mechanisms. Moreover, the termination of a task assigned to an agent may be compromised if the persistence of the agent itself is not ensured. In the case of a transactional mobile agent, we identify two services, Availability of the Sources and Atomic Commit, that can be supplied by more powerful entities located in a cloud. We propose a solution in which these two services are provided in a reliable and homogeneous way. To guarantee reliability, the proposed solution relies on a single agreement protocol that orders continuously all the new actions whatever the related transaction and service.

**DIONYSOS Project-Team**

# 5. New Results

## 5.1. Highlights of the Year

Pierre L'Ecuyer received the Award of Merit from the Canadian Operational Research Society, 2014.

We had one best paper award in 2014 on a novel architecture for resilient networks (see 5.8 ).
BEST PAPER AWARD :
[50] **IEEE International Conference on Innovations for Community Services**. D. LEQUÉRÉ, C. BETOULE, G. THOUENON, Y. HADJADJ-AOUL, A. KSENTINI, R. CLAVIER.

## 5.2. Quality of Experience

**Participants:** Yassine Hadjadj-Aoul, Adlen Ksentini, Gerardo Rubino, Bruno Sericola, Pantelis Frangoudis, César Viho, Quang Pham Tran Anh.

**PSQA.** We continue the development of the PSQA technology (Pseudo-Subjective Quality Assessment) in the area of Quality of Experience (QoE). PSQA is today a mature technology allowing to build measuring modules capable of quantifying the quality of a video or an audio sequence, as perceived by the user, when received through an IP network. It provides an accurate and efficiently computed evaluation of quality. Accuracy means that PSQA gives values close to those that can be obtained from a panel of human observers, under a controlled subjective testing experiment, following an appropriate standard (which depends on the type of sequence or application). Efficiency means that our measuring tool can work in real time. Observe that perceived quality is, in general, the main component of QoE when the application or service involves video and audio, or voice. PSQA works by analyzing the networking environment of the communication and some the technical characteristics of the latter. It works without any need to the original sequence (as such, it belongs to the family of *no-reference* techniques). It must be pointed out that a PSQA measuring or monitoring module is network-dependent and application-dependent. Basically, for each specific networking technology, and for any application or service, the module must be built from scratch. But once built, it works automatically and efficiently, allowing if necessary its use in real time, typically for controlling purposes.

**QoE and SLA.** On the applications side, we focused this year on using QoE estimates to drive service/application-level decisions. As a first use case, we proposed a multi-objective optimization framework for the problem of optimally selecting among a set of available hosting and network connectivity Service-Level Agreements (SLAs) for the migration of enterprise communication services (such as teleconferencing) to the Cloud [59]. Our framework captures the tradeoff between user experience and deployment cost, and offers a service provider the opportunity to weight these two conflicting criteria based on its preferences. Our approach is generic and can be applied to various application settings by appropriately selecting application-specific user experience models. For example, for enterprise voice teleconferencing we used the E-model for estimating user experience under a specific selection of hosting and network SLAs and a specific amount of resources (virtual machines) to deploy.

**QoE and collaborative projects.** We then considered QoE-aware content delivery, targeting in particular an environment where web and multimedia content is disseminated by over-the-top (OTT) providers, but assuming a level of cooperation between the content provider and the ISP (a trend which has started to become commonplace) [46]. We built on the outcome of our prior work [0], where we designed and implemented a network load estimation methodology and tool which operates by observing the delay behavior of the Precision Time Protocol (PTP) for network clock synchronization. After quantitatively establishing the link between network load and user experience, we proposed an architecture for OTT content delivery where user

[0] P.A. Frangoudis, A. Ksentini, Y. Hadjadj-Aoul, and G. Boime, "PTPv2-based network load estimation," Proc. IEEE ISPCS 2013. (This work was carried out in the context of the FUI project IPChronos, see 6.10 .)

requests are redirected to the data centers expected to offer optimal QoE, taking into account, among others, information about network load in the media path offered by our load estimation service (LES) in real time. In the same context, we developed a demonstrator where the LES is integrated as an additional network probe with the QoE monitoring architecture developed in the Celtic QuEEN project (see 7.2.1.1 ). Using a simple video QoE model which takes into account network load and video information (quality/resolution, bitrate), we implemented [0] an adaptation scheme for DASH video delivery which switches among video qualities based on QoE estimates received by the QuEEN software agent.

**QoE and PTPv2.** In [46], we make the case for an alternative use of the PTPv2 protocol: Adopting a learning approach, we observe its delay behavior during the protocol message exchange, derive models of its dependence on network load and build a real-time load estimation service. Then, as an application scenario of this service, we turn our attention to the provision of Over-the-Top (OTT) services. In such an environment, and assuming a level of cooperation between the ISP and the OTT provider, we demonstrate how our service can be used for estimating the QoE for web applications. To this end, we establish quantitatively the link between network load and user experience using a state-of-the-art web QoE monitoring framework, and show how our PTPv2-based load estimation scheme can integrated in an OTT service architecture and be utilized for load-aware, QoE-optimized content delivery decisions.

**QoE and reneging.** We consider in [45] an important Quality of Experience (QoE) indicator in mobile networks that is reneging of users due to impatience. We specifically consider a cell under heavy load conditions and compute the reneging probability by using a fluid limit analysis. By solving the fixed point equation, we obtain a new QoE perturbation metric quantifying the impact of reneging on the performance of the system. This metric is then used to devise a new pricing scheme accounting for reneging. We specifically propose several flavors of this pricing around the idea of having a flat rate for accessing the network and an elastic price related to the level of QoE perturbation induced by the communications.

**QoE-aware OLSR for Video Streaming over Wireless Multihop Networks.** Multi-hop environments can impact significantly ad-hoc network performance. In [57], we propose a routing algorithm based on optimized link state routing (OLSR), aimed at guaranteeing the quality of experience (QoE) of users in these types of networks. PSQA (see above in this same section) is used to estimate a mean opinion score (MOS), and then this MOS value is exploited by the source for selecting the appropriate path in the network. Moreover, an event- triggered based on the MOS value is used to provide more relevant information in selecting the best path by the source. The performance of this proposed mechanism was validated through intensive simulation under different scenarios. The results in [57] show that the proposed scheme outperforms other OLSR-based routing protocols particularly in a heavy load and high mobility scenario.

**QoE-Aware Routing for Video Streaming over VANETs.** In-vehicle multimedia applications are gaining interest since recent years. However, the high loss rate caused by high mobility in vehicular networks (VANETs) imposes several challenges in multimedia transmission. Moreover, in the context of multimedia, the quality of service (QoS)-based approaches assess the quality of streaming services through network-oriented metrics while the concept of quality of experience (QoE) is built upon the perception of users. In [58], a QoE-based routing protocol for video streaming over VANETs is proposed. By taking the mean opinion score (MOS) into account for path selection, good performance levels can be achieved, as shown by our simulation results.

## 5.3. Analytic models

**Participants:** Bruno Sericola, Gerardo Rubino, Raymond Marie.

**New book about Dependability Theory.** Dependability metrics are omnipresent in every engineering field, from simple ones through to more complex measures combining performance and dependability aspects of systems. The new book [69] written in the team, entitled “Markov Chains and Dependability Theory” and published in 2014 by Cambridge University Press (see also http://www.amazon.fr/Markov-Chains-Dependability-Theory-Gerardo/dp/1107007577/), presents the mathematical basis of the analysis of

[0] Our video adaptation scheme is implemented in the VLC open-source media player.

these metrics. The modelling context corresponds to the most used framework, Markov models. The book describes both basic results and specialised techniques. The authors first present discrete and continuous time Markov chains before focusing on dependability measures, which necessitate the study of Markov chains on a subset of states representing different user satisfaction levels for the modelled system. Topics covered include Markovian state lumping, analysis of sojourns on subset of states of Markov chains, analysis of most dependability metrics, fundamentals of performability analysis, and bounding and simulation techniques designed to evaluate dependability measures. As stated in its abstract, the book is of interest to graduate students and researchers in all areas of engineering where the concepts of lifetime, repair duration, availability, reliability and risk are important.

**Fluid models.** In [77] we study congestion periods in a finite fluid buffer when the input rate depends upon a recurrent Markov process; congestion occurs when the buffer content is equal to the buffer capacity. We consider the duration of congestion periods as well as the associated volume of lost information. We derive their distributions in a typical stationary busy period of the buffer. Our goal is to compute the exact expression of the loss probability in the system, which is usually approximated by the probability that the occupancy of the infinite buffer is greater than the buffer capacity under consideration. Moreover, by using general results of the theory of Markovian arrival processes, we show that the duration of congestion and the volume of lost information have phase-type distributions.

**Industrial Logistic Aspects.** Motivated by the consideration of clauses of penalty, we worked again on the determination of the probability distributions of the delays of unavailability of systems on the operational sites. By considering in particular a given type of spare, we show the important role played by the possible waiting time of the change during the occurrence of a breakdown. In particular we verify that the cumulative probability distribution of the delay of unavailability possesses a relatively low tail diminution as well as a high square of cœfficient of variation. Upper and lower bounds are highlighted in the simplest case. These results allow to calculate the risk inferred by the use of clauses of penalty; for example, by proposing an expression of the expectation of the cost of penalty imposed by unit of time if any unavailability exceeding a certain threshold is penalized [62]. If the possible waiting time of the change is the obsession of the specialists of the maintenance, the consideration of stock shortages in supply chains is often underestimated when these events are rare events. A related work consisted in showing that a low probability of break can be associated with a high coefficient of variation can have a very significant consequence [54].

We also studied the extension of our analytical method of calculation of the operational availability of a fleet of consequent systems deployed on a site and maintained by exchanges on the site of subsets (the LRU for *line repaired unit*) in the specific case where a policy of cannibalization is implemented. We propose an approximated method which is particularly adapted to the case of systems with strong operational availability because in this case the error inferred by the approximation remains low. The developed method consists in determining the expectation of the number of blocked systems due to the lack of change, in the presence of a policy of cannibalization. This expectation is directly associated with a loss of operational availability. At present, in the presence of a policy of cannibalization, the proposed solution concerns only the systems constituted by a series of LRU but the policy of cannibalization can be applied to all or part of the types of LRU [63].

## 5.4. Performance Evaluation

**Participants:** Pierre L'Ecuyer, Bruno Sericola, Romaric Ludinard.

**Network Monitoring and Fault Detection.** Monitoring a system consists in collecting and analyzing relevant information provided by the monitored devices, so as to be continuously aware of the system state (situational awareness). However, the ever growing complexity and scale of systems makes both real time monitoring and fault detection a quite tedious task. Thus the usually adopted option is to focus solely on a subset of information states, so as to provide coarse-grained indicators. As a consequence, detecting isolated failures or anomalies is a quite challenging issue. We propose in [39], [61] to address this issue by pushing the monitoring task at the edge of the network. We present a peer-to-peer based architecture, which enables nodes to adaptively and efficiently self-organize according to their "health" indicators. By exploiting both temporal and spatial

correlations that exist between a device and its vicinity, our approach guarantees that only isolated anomalies (an anomaly is isolated if it impacts solely a monitored device) are reported on the fly to the network operator. We show that the end-to-end detection process, *i.e.*, from the local detection to the management operator reporting, requires a logarithmic number of messages in the size of the network.

**Robustness Analysis of Large Scale Distributed Systems.** In the continuation of [81] which proposed an in-depth study of the dynamicity and robustness properties of large-scale distributed systems, we analyze in [13], the behavior of a stochastic system composed of several identically distributed, but non independent, discrete-time absorbing Markov chains competing at each instant for a transition. The competition consists in determining at each instant, using a given probability distribution, the only Markov chain allowed to make a transition. We analyze the first time at which one of the Markov chains reaches its absorbing state. When the number of Markov chains goes to infinity, we analyze the asymptotic behavior of the system for an arbitrary probability mass function governing the competition. We give conditions for the existence of the asymptotic distribution and we show how these results apply to cluster-based distributed systems when the competition between the Markov chains is handled by using a geometric distribution.

**Detection of distributed deny of service attacks** A Deny of Service (DoS) attack tries to progressively take down an Internet resource by flooding it with more requests than it is capable to handle. A Distributed Deny of Service (DDoS) attack is a DoS attack triggered by thousands of machines that have been infected by a malicious software, with as immediate consequence the total shut down of targeted web resources (*e.g.*, e-commerce websites). A solution to detect and to mitigate DDoS attacks it to monitor network traffic at routers and to look for highly frequent signatures that might suggest ongoing attacks. A recent strategy followed by the attackers is to hide their massive flow of requests over a multitude of routes, so that locally, these flows do not appear as frequent, while globally they represent a significant portion of the network traffic. The term "iceberg" has been recently introduced to describe such an attack as only a very small part of the iceberg can be observed from each single router. The approach adopted to defend against such new attacks is to rely on multiple routers that locally monitor their network traffic, and upon detection of potential icebergs, to inform a monitoring server that aggregates all the monitored information to accurately detect icebergs. Now, to prevent the server from being overloaded by all the monitored information, routers continuously keep track of the $c$ (among $n$) most recent high flows (called items) prior to sending them to the server, and throw away all the items that appear with a small probability. Parameter $c$ is dimensioned so that the frequency at which all the routers send their $c$ last frequent items is low enough to enable the server to aggregate all of them and to trigger a DDoS alarm when needed. This amounts to compute the time needed to collect $c$ distinct items among $n$ frequent ones. A thorough analysis of the time needed to collect $c$ distinct items appears in [71].

**Randomized Message-Passing Test-and-Set.** In [74], we present a solution to the well-known Test&Set operation in an asynchronous system prone to process crashes. Test&Set is a synchronization operation that, when invoked by a set of processes, returns yes to a unique process and returns no to all the others. Recently, many advances in implementing Test&Set objects have been achieved, but all of them target the shared memory model. In this paper we propose an implementation of a Test&Set object in the message passing model. This implementation can be invoked by any number $p \leq n$ of processes where $n$ is the total number of processes in the system. It has an expected individual step complexity in $O(\log p)$ against an oblivious adversary, and an expected individual message complexity in $O(n)$. The proposed Test&Set object is built atop a new basic building block, called selector, that allows to select a winning group among two groups of processes. We propose a message-passing implementation of the selector whose step complexity is constant. We are not aware of any other implementation of the Test&Set operation in the message passing model.

**Call centers.** We develop research activities around the analysis and design of call centers, from a performance perspective. In [56], we focus on the scheduling problem (which task must be done by which worker at each period of time). We show that a Constraint Programming model can be used to solve large instances of this type of optimization work. In [21], we study call routing policies for call centers with multiple call types and multiple agent groups, focusing on the case of small and medium size centers, whose behavior may differ from those obtained in heavy-traffic regimes, and for which non-work-conserving policies can perform better. We

propose a routing policy based on weights, expressed as linear functions of the call waiting times and agent idle times, or number of idle agents, following a simulation-based optimization approach.

## 5.5. Network Economics

**Participants:** Bruno Tuffin, Pierre L'Ecuyer.

The general field of network economics, analyzing the relationships between all acts of the digital economy, has been an important subject for years in the team. The whole problem of network economics, from theory to practice, describing all issues and challenges, is described in our book [67].

Among the topics we have particularly focused on, the network neutrality debate was a major concern in 2014. In the position paper [79], Bruno Tuffin and his co-author Patrick Maillé discuss for a large audience the issues and challenges of network neutrality in response to the European parliament text voted in April 2014. A related (and often forgotten) issue, the recently raised search neutrality debate questions the ranking methods implemented by search engines: when a search is performed, do they (or should they) display the web pages ordered according to the quality-of-experience (relevance) of the content? In [22], we analyze that question in a setting when content is offered for free, content providers making revenue through advertising. For content providers, determining the amount of advertising to add to their content is a crucial strategic decision. Modeling the trade-off between the revenue per visit and the attractiveness, we investigate the interactions among competing content providers as a non-cooperative game, and consider the equilibrium situations to compare the different ranking policies. Our results indicate that when the search engine is not involved with any high-quality content provider, then it is in its best interest to implement a neutral ranking, which also maximizes user perceived quality-of-experience and favors innovation. On the other hand, if the search engine controls some high-quality content, then favoring it in its ranking and adding more advertisement yields a larger revenue. This is not necessarily at the expense of user perceived quality, but drastically reduces the advertising revenues of the other content providers, hence reducing their chances to innovate.

But while ISPs and search engines are almost the only Internet actors being pointed out as potentially non neutral, we investigate the economic impact and strategies of Content Delivery Networks (CDNs), Internet actors that reduce the capacity needs in the backbone network and improve the quality perceived by users. The growing importance of Content Delivery Network (CDN) in the value chain of content delivery raises concerns about the neutrality of these players. We consider in [52] the so-called push and pull models where the traffic is paid by the sender or the receiver, respectively, as well as the situation where the CDN is (vertically) integrated to, i.e., owned by, an Internet Service Provider (ISP). We then discuss the implication of CDNs into the network neutrality debate, another issue forgotten by researchers and regulators. We also propose in [53] a model to analyze the impact of revenue-oriented CDN management policies on the fairness of the competition among two content providers that use CDN services to deliver contents. We show that there exists a unique optimal revenue maximizing policy for a CDN actor –the dimensioning and allocation of its storage capacity– that depends on prices for service/transport/storage, and on the distribution of content popularity. Using data from the analysis of traces from two major content providers (YouTube Live and justin.tv), we remark that a CDN remains a relatively neutral actor even when one of the content providers it serves tries to monopolize the CDN storage space by implementing an aggressive policy to harm its competitors.

Finally, when a customer searches for a keyword at a classified ads website, at an online retailer, or at a search engine (SE), the platform has exponentially many choices in how to sort the output to the query. The two extremes are (a) to consider a ranking based on relevance only, which attracts more customers in the long run because of perceived quality, and (b) to consider a ranking based on the expected revenue to be generated by immediate conversions, which maximizes short-term revenue. Typically, these two objectives are not perfectly positively correlated and hence the main question is what middle ground between them should be chosen. We introduce in [78] stochastic models and propose effective solution methods that can be used to optimize the ranking considering long-term revenues. A key feature of our model is that customers are quality-sensitive and are attracted to the platform or driven away depending on the average relevance of the output. The proposed methods are of crucial importance in e-business and encompass: (i) classified ad websites which can favor paid ads by ranking them higher, (ii) online retailers which can rank products they sell according to buyers'

interests and/or the margins these products have, (iii) SEs which can position the content that they serve higher in the output page than third-party content to keep users in their platforms for longer and earn more. This goes in detriment of just offering rankings based on relevance only and is directly linked to the current search neutrality debate.

## 5.6. Monte Carlo

**Participants:** Bruno Tuffin, Gerardo Rubino, Pierre L'Ecuyer.

We maintain a research activity in different areas related to dependability, performability and vulnerability analysis of communication systems, using both the Monte Carlo and the Quasi-Monte Carlo approaches to evaluate the relevant metrics. Monte Carlo (and Quasi-Monte Carlo) methods often represent the only tool able to solve complex problems of these types. A review of Monte Carlo, Quasi-Monte Carlo and pseudo-random generation can be found in [66]. In [27], we examine some properties of the points produced by certain classes of long-period linear multiple recursive random number generators. These generators have their parameters selected in special ways to make the implementation faster. We show that as a result, the points produced by these generators have a poor lattice structure, and a poor initialization of the state can have long-lasting impact, because of the limited diffusion capacity of the recurrence.

However, when the events of interest are rare, simulation requires a special attention, to accelerate the occurrence of the event and get unbiased estimators of the event of interest with a sufficiently small relative variance. This is the main problem in the area. Dionysos' work focuses then on dealing with the rare event situation. In [20], we present several state-of-the-art Monte Carlo methods for simulating and estimating rare events. Among variance reduction methods, the most prominent ones for this purpose are Importance Sampling (IS) and Multilevel Splitting, also known as Subset Simulation. Some recent results on both aspects are described, motivated by theoretical issues as well as by applied problems.

A non-negligible part of our activity on the application of rare event simulation was about the evaluation of static network reliability models, with links subject to failures. Exact evaluation of static network reliability parameters belongs to the NP-hard family and Monte Carlo simulation is therefore a relevant tool to provide their estimations. In [34], we propose an adaptive parameterized method to approximate the zero-variance change of measure. The method uses two rough approximations of the unreliability function, conditional on the states of any subset of links being fixed. One of these approximations, based on mincuts, under-estimates the true unknown unreliability, whereas the other one, based on minpaths, over-estimates it. Our proposed change of measure takes a convex linear combination of the two, estimates the optimal (graph-dependent) coefficient in this combination from pilot runs, and uses the resulting conditional unreliability approximation at each step of a dynamic importance sampling algorithm. This new scheme is more general and more flexible than a previously-proposed zero-variance approximation one, based on mincuts only, and which was shown to be robust asymptotically when unreliabilities of individual links decrease toward zero. Our numerical examples show that the new scheme is often more efficient when low unreliability comes from a large number of possible paths connecting the considered nodes rather than from small failure probabilities of the links. Another paper, reference [18], focuses on another technique, known as Recursive Variance Reduction (RVR) estimator which approaches the unreliability by recursively reducing the graph from the random choice of the first working link on selected cuts. This previously known method is shown to not verify the bounded relative error (BRE) property as reliability of individual links goes to one, i.e., the estimator is not robust in general to high reliability of links. We then propose to use the decomposition ideas of the RVR estimator in conjunction with the IS technique. Two new estimators are presented: the first one, called Balanced Recursive Decomposition estimator, chooses the first working link on cuts uniformly, while the second, called Zero-Variance Approximation Recursive Decomposition estimator, combines RVR and our zero-variance IS approximation. We show that in both cases BRE property is verified and, moreover, that a vanishing relative error (VRE) property can be obtained for the Zero-Variance Approximation RVR under specific sufficient conditions. A numerical illustration of the power of the methods is provided on several benchmark networks. Continuing the analysis of existing method, we have described in [44] a necessary and sufficient condition for a well known technique called Fishman's method to verify BRE and have realized a deep analysis of the technique.

But in the literature and the previously described static network reliability models one typically assumes that the failures of the components of the network are independent.This simplifying assumption makes it possible to estimate the network reliability efficiently via specialized Monte Carlo algorithms. Hence, a natural question to consider is whether this independence assumption can be relaxed, while still attaining an elegant and tractable model that permits an efficient Monte Carlo algorithm for unreliability estimation. In [75], we provide one possible answer by considering a static network reliability model with dependent link failures, based on a Marshall-Olkin copula, which models the dependence via shocks that take down subsets of components at exponential times, and propose a collection of adapted versions of permutation Monte Carlo (PMC, a conditional Monte Carlo method), its refinement called the turnip method, and generalized splitting (GS) methods, to estimate very small unreliabilities accurately under this model. The PMC and turnip estimators have bounded relative error when the network topology is fixed while the link failure probabilities converge to 0. When the network (or the number of shocks) becomes too large, PMC and turnip eventually fail, but GS works nicely for very large networks, with over 5000 shocks in our examples. [65] focuses on the application of our zero-variance approximation IS estimator to this same type of model.

Another family of models of interest in the group are the highly reliable Markovian systems, made of components subject to failures and repairs. We describe in [60] how importance sampling can be applied to efficiently estimate the average interval availability of those models. We provide a methodology for approximating the zero-variance change of measure. The method is illustrated to be very efficient on a small example, compared with standard importance sampling strategies developed in the literature.

Finally, in Quasi-Monte Carlo (QMC), the error when estimating an integral uses a deterministic sequence (instead of a random one) called a low discrepancy sequence and having the property to spread quickly over the integration domain. The estimation error is bounded by the product of a quantity depending on the discrepancy of the sequence and the variation of the integrand. But this bound is proved to be useless in practice. By combining MC and QMC methods, we can benefit from the advantages of both approaches: error estimation from MC and convergence speed from QMC. Randomized quasi-Monte Carlo (RQMC) is another class of methods for reducing the noise of simulation estimators, by sampling more evenly than with standard MC. In [37], we analyze the convergence rate of the *array-RQMC* technique, a randomized QMC method we have previously designed and devoted to the simulation of Markov chains.

In [19], we propose a method for estimating performability metrics built upon non-binary network states, determined by the hop distances between distinguished nodes. In other words, we explore the analysis of a generalization of network reliability, particularly relevant for instance in telecommunications. The estimation is performed by a Monte Carlo simulation method where the sampling space is reduced using edge sets known as $d$-pathsets and $d$-cutsets. Numerical experiments over two mesh-like networks are presented. They show significant efficiency improvements relative to the crude Monte Carlo method, in particular as link failures become rare events, which is usually the case in most real communication networks.

## 5.7. Wireless Networks

**Participants:** Osama Arouk, Btissam Er-Rahmadi, Adlen Ksentini, Yassine Hadjadj-Aoul, Quang Pham Tran Anh, Hyunhee Park, César Viho.

We continue our activities around wireless and mobile networks, where we focus particularly on 4G/5G networks as well as on a new mobile architecture known as mobile cloud.

**LTE improvements**. In [35], we investigated, at both the core network (EPC) and Radio Access Network (RAN), the impact of caching the shared content among users. We reviewed the different locations were data could be cached and their impacts on user QoS/QoE. In [33], we proposed several new mechanisms to handle the gateway relocation in the context of highly decentralized mobile network. To evaluate these mechanisms, we proposed an analytical model based on Markov Chains, whereby we captured the randomness of user mobility and its impact on the user QoS in terms of the probability to be connected to the optimal gateway, the drop rate, etc. In [32], we devised an agile admission control mechanism that anticipates QoS/QoE degradation and proactively defines policies for admitting UEs handing-in from the macro network to the

small cell network. It also enables IP flow mobility between small cells and macro networks. We provided an analytical model to the admission control mechanism based on Markov Decision Processes (MDP). The ultimate objective of the proposed model is to derive the optimal policy (i.e., reject or accept flows in the macro or the small cell) which maximizes users' QoE under different load scenarios (low and high load user traffic). Another work regarding small cells in LTE was proposed in [76], where we used the small cell principle to extend the mobile network coverage in emerging countries that not include a wired infrastructure. The proposed framework aims to backhaul the small cell with the less costly connection, while ensuring minimal QoS to users. In this vein, we formulated this problem through an Integer Linear Program (ILP), and solve it for small network sizes. For large instances of the network size, we proposed two new heuristics. In [30], we investigated network decentralization in conjunction with the Selective IP Traffic Offload (SIPTO) approaches to handle the mobile increased data traffic. We first devised different approaches based on a per destination domain name basis, which offer operators a fine-grained control to determine whether a new IP connection should be offloaded or accommodated via the core network. Two of our solutions are based on Network Address Translation (NAT) named simple-NATing and twice-NATing, while a third one employs simple tunneling and a forth proposal adopts multiple Access Point Names (APNs). We also proposed methods enabling User Equipment (UEs), both in idle and active modes and while being on the move, to always have efficient Packet Data Network (PDN) connections. A qualitative analysis and a simulation study compared the different approaches with respect to cost, complexity, service continuity and network performance, demonstrating the significance of the proposed schemes for multimedia applications.

**M2M**. We addressed another type of traffic that appeared these last years, namely Machine to Machine (M2M) communication or Machine Type Communication (MTC). Such traffic is known by its intensity and its impact on increasing congestion in both parts of 4G networks, the Radio Access Network (RAN) and the core network. The main spirit of the proposed solutions is to proactively anticipate system overload by reducing the amount of MTC signaling messages exchanged in normal network operations. In [49] we introduced a solution that operates at the core network. We proposed that the Mobility Management Entity (MME), or an alike core network node, computes the device trigger rate that alleviates congestion, and communicates this value to the MTC-Interworking Function (MTC-IWF) element that enforces MTC traffic control, via admission control or data aggregation, on the device trigger request rate received from the different MTC servers.

As mentioned earlier, the MTC would impact not only the EPC part, but also the RAN. Group paging is currently considered as one of the most efficient mechanisms proposed to alleviate the problem of the RAN overload. In [42] we introduced a new solution to improve the performance of the current group paging method and overcome its disadvantages. The proposed solution is intended for MTC devices in connected mode state, in which they have an RRC context without being synchronized with the network. In [41] we devised a novel algorithm which estimates the network status (the number of active devices), thus better controlling the RAN access. Unlike most existing methods that consider only one channel, the proposed solution uses the statistics of all the channels in order to estimate the number of arrivals (UE and MTC devices) in each RA (Random Access) slot.

Most of the above-proposed solutions are basically incremental ones. In [31], we devised a complete new architectural vision to support MTC in mobile networks. This vision relies on the marriage of mobile networks and the cloud, specifically based on Network Function Virtualization (NFV). The proposed solution simplifies the network attach procedure for MTC devices by creating only one NFV MTC function that groups all the usual procedures. By doing so, the proposed solution is able to create and scale instances of NFV MTC functions on demand and in an elastic manner to cope with any sudden increase in traffic generated by MTC devices.

**Wireless Sensor Networks (WSN)**. WSNs are complex systems that are mainly limited by the battery life of the nodes in order to have an adequate performance. In most cases, it is possible to have a re-deployment of new nodes in order to prolong the systems lifetime. This leads to a situation where some nodes have a low energy level while other nodes (the majority of nodes a few instants after the re-deployment procedure) have high energy levels. In these environments, it is clear that ancient nodes, those with low energy levels, have to contend for the shared medium against the majority of high energy nodes. As such, the remaining

battery life of low energy nodes would be rapidly consumed. In [64], we propose to extend the battery life of low energy nodes by means of assigning prioritized access to the shared channel to those nodes. The goal is to content among a low population of such nodes, while delaying the contention access of high energy nodes which can support higher number of collisions before energy depletion. This is done by studying two different transmission strategies referred to as "hard" and "soft" transmission probabilities. Results show that a soft transmission strategy achieves better results in terms of reduced energy consumption than both the conventional protocol or a hard transmission assignment.

The communication between nodes is the greedy factor to the energy consumption. One important mechanism to reduce the energy consumption is the in-network data aggregation. This mechanism removes repeated and unnecessary data readings and thus cuts on the energy used in communications. In [14] we reviewed the state of art on this topic. Then, we proposed a classification of the available solutions according to the way the aggregation is done. In [15], we addressed the reliable minimum data aggregation scheduling problem in wireless sensor networks under multi-channel frequency use. The proposed solution ensures the collection reliability and reduces the latency in disseminating aggregated data to the base-station over multi-frequency radio links. Another mechanism to improve energy efficiency is to optimize link scheduling when using TDMA-based techniques and data fragmentation when using slotted CSMA/CA access methods. In this line, we proposed a protocol, named DLSP, with the objective of achieving both low energy consumption and low latency in Wireless Sensor Networks. DLSP takes advantage of the spatial reuse of interference-free time slots by means of conflicts graphs. Unlike the previous studies that often consider saturated nodes, we propose to relax the saturation assumption in order to maintain good performance when some of the nodes have no data to send citemouloud:hal-01101396. In [55], we noticed that the standardized slotted CSMA/CA may lead to a wastage of the bandwidth utilization and an additional transmission delay. This drawback is mainly caused by Deferred Transmission in the CSMA/CA algorithm at the end of the superframe, when there is not sufficient time to complete the frame transmission. Thus, we proposed to fragment a data frame into a short frame and attempt its transmission in the current frame and transmit the remaining frame in the next superframe. The data fragmentation mechanism was modeled using a Markov chain. A non-saturated traffic and acknowledgement transmission are considered in our analysis.

**High data rate WiFi networks**. The IEEE 802.11ac Task Group (TGac) is actively working on an amendment that allows WLAN to reach a maximum aggregate network throughput up to 7 Gbps on bands below 6 GHz. In particular, the standard envisions a maximum Medium Access Control (MAC) throughput of at least 500 Mbps for a single user, and at least 1 Gbps in case of multiple users. In [36] we proposed an analysis of the IEEE 802.11ac TXOP Sharing mechanism, which was recently introduced by the 802.11ac group, by providing a Markov chain-based model. Based on the proposed Markov chain, we provided an analytical model of the achievable throughput for each AC. Accordingly, we can analyze the impact of the TXOP Sharing on the throughput of each AC, hence highlighting the improvement achieved in terms of bandwidth utilization and channel access fairness among the different ACs.

**Mobile cloud**. One of the 5G-architecture visions considers the usage of clouds to build mobile networks and help in decentralizing mobile networks on demand, elastically, and in the most cost-efficient way. This concept of carrier cloud becomes of vital importance knowing that several cloud providers are distributing their cloud/network, globally deploying more regional data centers, to meet their ever-increasing business demands. As an important enabler of the carrier cloud concept, network function virtualization (NFV) is gaining great momentum among industries. NFV aims for decoupling the software part from the hardware part of a carrier network node, traditionally referring to a dedicated hardware, single service and single-tenant box, that is using virtual hardware abstraction. Network functions become thus a mere code, runnable on a particular, preferably any, operating system and on top of a dedicated hardware platform. The ultimate objective is to run network functions as software in standard virtual machines (VMs) on top of a virtualization platform in a general-purpose multi-service multi-tenant node (e.g., Carrier Grade Blade Server) put into the cloud. In [26], we presented a LISP-based implementation of the Follow Me Cloud (FMC) concept, whereby mobile services hosted in federated clouds follow mobile users as they move and according to their needs. This implementation clearly demonstrates the feasibility of the FMC concept. On the other hand, service migration in FMC may be an expensive operation given the incurred cost in terms of signaling messages and

data transferred between DCs. Indeed, decision on service migration defines therefore a tradeoff between cost and user perceived quality. In [48] we addressed this tradeoff by modeling the service migration procedure using a Markov Decision Process (MDP). The aim was to formulate a decision policy that determines whether to migrate a service or not when the concerned User Equipment (UE) is at a certain distance from the source DC.

In order to meet the general needs of mobile operators, efficient mobile cloud must give high importance to the placement/instantiation of mobile network functions (such as data anchor gateways) in the federated cloud. In [43] we argued the need of using service/application type and requirements as metrics for efficiently: (i) create virtual instance of the Packet Data Network Gateway (PDN-GW); (ii) select the virtual PDN-GW for UEs with specific application type. After modeling this procedure though a nonlinear Optimization Problem (OP) and proving it as a NP-hard problem, we proposed three solutions to solve this issue.

**Wireless Local Area Networks.** User-centric networking has emerged as a disruptive new communication paradigm. We particularly focused on its expressions in *wireless* networking and the challenges it brings about [23]. In this context, by means of testbed experiments and simple analytic models, we quantified the upper bounds on VoIP capacity of a purely user-centric secure VoIP communications scheme that we designed, identifying the major quality degradation factors. Our results have shown that typical user Wi-Fi equipment can sustain a satisfactory number of concurrent secure VoIP sessions with acceptable QoE and, at the same time, protection from malicious user activity can be offered to access providers, while a level of roaming privacy can be guaranteed [24]. We then studied the role of users in wireless network management tasks. In particular, we proposed a scheme where monitoring the topology of Wi-Fi deployments is crowdsourced to roaming users, who submit reports on wireless coverage in their vicinity [25]. Topology information can then be used as input to reconfiguration mechanisms, such as channel assignment schemes. Users cannot be assumed trustworthy, though. They can engage in fraudulent reporting, which, unless specific countermeasures are in place, can severely impact one's view of the network topology. To this end, we designed and implemented an architecture for accurate Wi-Fi topology discovery, devising a reputation-based mechanism to tackle realistic and simple to implement attacks. We have shown analytically and via simulation that, even in the presence of large numbers of attackers, our user-centric scheme significantly outperforms pure infrastructure-based approaches, where monitoring is carried out only by trusted Access Points.

In another line of research, we focused on efficiently integrating wireless users in an Information-Centric Network (ICN) architecture. In ICN, multicast content delivery is the norm. At the same time, wireless multicast is problematic. To address this issue, we took advantage of the content awareness inherent in ICN and proposed a relay-based approach for local wireless multicasting: ICN information *scoping* mechanisms assist in expressing content semantics and, in turn, encoding the heterogeneous performance requirements of different content/application types. Under this premise, we proposed a multiobjective optimization approach for relay selection and multicast transmission rate assignment which allows to optimize for reliability, delivery time, or energy cost on a per content basis [47].

**Energy saving**. Another part of our activities in wireless network are related to energy saving. Indeed, one of the biggest problem today in the wireless world is that wireless devices are battery-driven, which reduce their operating lifetime. The experimental measurements we have achieved in [16] and [17] revealed that operating system overhead causes a drop in performance and energy consumption properties as compared to the GPP in case of certain low video qualities. We propose, thus, a new approach for energy-aware processor switching (GPP or DSP) which takes into consideration video quality. We show the pertinence of our solution in the context of adaptive video decoding and implement it on an embedded Linux operating system.

**Adaptive Beam Scheduling for Scalable Video Multicast in Wireless Networks.** Design of efficient multicast for a scalable video coding (SVC) streaming combined with directional beamforming is a challenging issue. In [29], we propose a QoE-aware directional beam scheduling (QBS) scheme which optimizes overall quality of experience (QoE) for multirate multicast of SVC, with beamforming in wireless networks. We optimally schedule different SVC layers to different beams and rate modulations. We provide a mixed integer linear programing (MILP) formulation of the problem, and then propose a heuristic algorithm. Extensive sim-

ulation results demonstrate that QBS can increase the overall QoE and can satisfy a minimum expected QoE for all users.

## 5.8. Future networks and architectures

**Participants:** Damien Le Quéré, Adlen Ksentini, Yassine Hadjadj-Aoul, Jean-Michel Sanner.

**LOCARN**. LOCARN (i.e. Low Opex & Capex Architecture for Resilient Networks) is a flat, dynamic and very simple packet architecture that focuses on plug-and-play guidance to provide flexibility and resiliency on the transport of client data traffics. To that end, the counterpart of the solution is a significant overhead due to the generation of control plane packets. In [50], we proved that in typical meshed operators transport networks applications, (i.e. infrastructures having high data-rates and high resiliency requirements), the LOCARN overhead is acceptable up to thousands of communications. In [51], we introduced two proposals that permit to increase the amount of simultaneous communications while maintaining the good properties of the initial design.

**SDN**. We started an activity on Software Defined Networking (SDN), a recent idea proposed to handle network management problems. SDN are becoming an important issue with the ever-increasing network complexity. They are proposed as an alternative to the current architecture of the Internet, which cannot meet the supported services requirements such as Quality of Service/Experience (Qos/QoE), security and energy consumption. We particularly address the scalability issue by proposing an automated hierarchical controller-based architecture handling the whole control chain.

**DIVERSE Project-Team**

# 6. New Results

## 6.1. Highlights of the Year

**"Globalizing Modeling Languages" appears in IEEE Computer Magazine.** This paper synthesizes our vision of how domain-specific languages form the foundations of global software development. Its appearance in a highly visible venue is major milestone for the dissemination and impact of our work about the diversity of languages.

**DiverSE extremely present at the SPLC conference.** SPLC is the main international conference for software product line engineering. In 2014, the DiverSE team had a very strong presence at this conference, presenting novel scientific contributions, results of industrial collaborations, and demonstrations of latest software tools.

## 6.2. Results on 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 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. In [56] we present the overall vision that we develop in the DiverSE team about Software Language Engineering. The main results on this topic are presented below.

### *6.2.1. Globalization of Domain Specific Languages*

In the software and systems modeling community, research on domain-specific modeling languages (DSMLs) focuses on technologies for developing languages and tools to increase the effectiveness of domain experts. Yet, there is a lack of support to explicitly relate concepts expressed in different DSMLs, which prevents software and system engineers to reason about information spread across models describing different system aspects. Supporting coordinated use of DSMLs leads to what we call the globalization of modeling languages [20]. In such a context, we develop a research initiative that broadens the DSML research focus beyond the development of independent DSMLs to one that supports globalized DSMLs, that is, DSMLs that facilitate coordination of work across different domains of expertise. We also provid a formal framework to prove the correctness of model driven engineering composition operators [57].

### *6.2.2. Meta-Language for the Concurrency Concern in DSLs*

Concurrency is of primary interest in the development of complex software-intensive systems, as well as the deployment on modern platforms. However, reifying the definition of the DSL concurrency remains a challenge. This hinders: a) the development of a complete understanding of the DSL semantics; b) the effectiveness of concurrency-aware analysis techniques; c) the analysis of the deployment on parallel architectures. In this context, we present MoCCML, a dedicated meta-language for formally specifying the concurrency concern within the definition of a DSL [44]. The concurrency constraints can reflect the knowledge in a particular domain, but also the constraints of a particular platform. MoCCML comes with a complete language workbench to help a DSL designer in the definition of the concurrency directly within the concepts of theDSL itself, and a generic workbench to simulate and analyze any model conforming to this DSL. Mo CCML is illustrated on the definition of an lightweight extension of SDF (SynchronousData Flow).

### 6.2.3. Automating Variability Model Inference for Component-Based Language Implementations

Componentized language frameworks, coupled with variability modeling, have the potential to bring language development to the masses, by simplifying the configuration of a new language from an existing set of reusable components. However, designing variability models for this purpose requires not only a good understanding of these frameworks and the way components interact, but also an adequate familiarity with the problem domain. In [68] we propose an approach to automatically infer a relevant variability model from a collection of already implemented language components, given a structured, but general representation of the domain. We describe techniques to assist users in achieving a better understanding of the relationships between language components, and find out which languages can be derived from them with respect to the given domain.

### 6.2.4. Metamorphic Domain-Specific Languages

External or internal domain-specific languages (DSLs) or (fluent) APIs? Whoever you are – a developer or a user of a DSL – you usually have to choose side; you should not! What about metamorphic DSLs that change their shape according to your needs? Our 4-years journey of providing the "right" support (in the domain of feature modeling), led us to develop an external DSL, different shapes of an internal API, and maintain all these languages. A key insight is that there is no one-size-fits-all solution or no clear superiority of a solution compared to another. On the contrary, we found that it does make sense to continue the maintenance of an external and internal DSL. Based on our experience and on an analysis of the DSL engineering field, the vision that we foresee for the future of software languages is their ability to be self-adaptable to the most appropriate shape (including the corresponding integrated development environment) according to a particular usage or task. We call metamorphic DSL such a language, able to change from one shape to another shape [27].

### 6.2.5. Adapting mutation testing for model transformations

Due to the specificities of models and transformations, classical software testing techniques have to be adapted. Among these techniques, mutation analysis has been ported and a set of mutation operators has been defined. However, mutation analysis currently requires a considerable manual work and is hampered by the test data set improvement activity. This activity is seen by testers as a difficult and time-consuming job, and reduces the benefits of the mutation analysis.

We provide a model transformation traceability mechanism, in conjunction with a model of mutation operators and a dedicated algorithm, to automatically or semi-automatically produce test models that detect new faults [18].

### 6.2.6. Efficient model cloning for analysis

We propose an original approach that exploits the fact that operations rarely modify a whole model. Given a set of immutable properties, our cloning approach determines the objects and fields that can be shared between the runtime representations of a model and its clones. Our generic cloning algorithm is parameterized with three strategies that establish a trade-off between memory savings and the ease of clone manipulation. We evaluated memory footprints and computation overheads with 100 randomly generated metamodels and models [40]. We have also drawn the research roadmap to exploit these efficient clone operations to analyze multidimensional execution traces [41].

## 6.3. Results on Variability Modeling and Engineering

### 6.3.1. Engineering Interactive Systems

In agreement with our permanent effort to validate the techniques we propose on real use cases in various domains, we applied seminal MDE to interactive systems engineering. This led to two collaborations. The first one has been conducted with 3D Collaborative Virtual Environments (3D CVE) researchers. Despite the increasing use of 3D CVE, their development is still a cumbersome task. The various concerns to consider

(distributed system, 3D graphics, *etc.*) complexify their development as well as their evolution. We propose to leverage MDE for developing 3D CVEs [45]. We have shown how a 3D CVE framework benefits from a DSL we built using state-of-the-art MDE technologies. The benefits are multiple: 3D CVEs designers can focus on the behavior of their virtual objects without bothering with distributed and graphics features; configuring the content of 3D CVEs and their deployment on various software and hardware platforms can be automated through code generation.

The second collaboration is international and has been conducted with software visualization researchers. Current metamodel editing tools are based on standard visualization and navigation features, such as physical zooms. However, as soon as metamodels become larger, navigating through large metamodels becomes a tedious task that hinders their understanding. In this work, we promote the use of model slicing techniques [102] to build visualization techniques dedicated to metamodels [37]. This approach is implemented in a metamodel visualizer, called `Explain`.

### 6.3.2. Variability management in regulatory requirements and system engineering

Nuclear power plants are some of the most sophisticated and complex energy systems ever designed. These systems perform safety critical functions and must conform to national safety institutions and international regulations. In many cases, regulatory documents provide very high level and ambiguous requirements that leave a wide margin for interpretation. As the French nuclear industry is now seeking to spread its activities outside France, it is but necessary to master the ins and the outs of the variability between countries safety culture and regulations. This sets both an industrial and a scientific challenge to introduce and propose a product line engineering approach to an unaware industry whose safety culture is made of interpretations, specificities, and exceptions. We have developed two contributions within the French R&D project CONNEXION, while introducing variability modeling to the French nuclear industry [66], [34].

As part of the VaryMDE project (a bilateral collaboration between Thales and Inria) we have developed techniques to generate counter-examples (also called anti-patterns) of model-based product lines [22]. The goal is to infer (1) guidelines or domain-specific rules to avoid earlier the specification of incorrect mappings (2) testing oracles for increasing the robustness of derivation engines given a modeling language. We have applied the approach in the context of a real industrial scenario with Thales involving a large-scale metamodel.

### 6.3.3. Handling testing challenges in product line engineering

Testing techniques in industry are not yet adapted for product line engineering (PLE).

We have developed original contributions to adapt model-based testing for PLE [65], [63], [13]. We equip usage models, a widely used formalism in MBT, with variability capabilities. Formal correspondences are established between a variability model, a set of functional requirements, and a usage model. An algorithm then exploits the traceability links to automatically derive a usage model variant from a desired set of selected features. The approach is integrated into the MBT tool MaTeLo and is currently used in industry.

We have also developed a variability-based testing approach to derive video sequence variants. The ideas of our VANE approach are i) to encode in a variability model what can vary within a video sequence; ii) to exploit the variability model to generate testable configurations; iii) to synthesize variants of video sequences corresponding to configurations. VANE computes T-wise covering sets while optimizing a function over attributes [50], [25].

### 6.3.4. Reverse engineering variability models

We have developed automated techniques and a comprehensive environment for synthesizing feature models from various kinds of artefacts (e.g. propositional formula, dependency graph, FMs or product comparison matrices). Specifically we have elaborated a support (through ranking lists, clusters, and logical heuristics) for choosing a sound and meaningful hierarchy [42]. We have performed an empirical evaluation on hundreds of feature models, coming from the SPLOT repository and Wikipedia [108]. We have showed that a hybrid approach mixing logical and ontological techniques outperforms state-of-the-art solutions (to appear in Empirical Software Engineering journal in 2015 [19]). Beyond the reverse engineering of variability, our work has numerous practical applications (e.g., merging multiple product lines, slicing a configuration process).

### 6.3.5. Product comparison matrices

Product Comparison Matrices (PCMs) constitute a rich source of data for comparing a set of related and competing products over numerous features. Despite their apparent simplicity, PCMs contain heterogeneous, ambiguous, uncontrolled and partial information that hinders their efficient exploitations. We have first elaborated our vision and identify research challenges for an exploitation of PCMs when engineering comparators, configurators, or other services [67].

We have formalized PCMs through model-based automated techniques and developed additional tooling to support the edition and re-engineering of PCMs [43]. 20 participants used our editor to evaluate our PCM metamodel and automated transformations. The empirical results over 75 PCMs from Wikipedia show that (1) a significant proportion of the formalization of PCMs can be automated: 93.11% of the 30061 cells are correctly formalized; (2) the rest of the formalization can be realized by using the editor and mapping cells to existing concepts of the metamodel.

The ASE'2014 paper opens avenues for engaging a community in the mining, re-engineering, edition, and exploitation of PCMs that now abound on the Internet. We have launched an open, collaborative initiative towards this direction http://www.opencompare.org

## 6.4. Results on Heterogeneous and dynamic software architectures

This year, we focused on the challenges that use *models@runtime* for resource-constrained and resource-aware systems. Our main results are in the following four subdomains:

- We designed an adaptive monitoring framework for component-based systems in which we highlight the benefits of using models@runtime for adaptive monitoring.
- We improved models@runtime technologies for resource-constrained devices.
- We designed efficient reasoning techniques for dynamic software architecture, focusing in particular on resource consumption optimization challenges.
- We performed several experiments on the Internet of Things application domain.

The next section details our experiments.

### 6.4.1. Resource-aware dynamic architecture

Modern component frameworks support continuous deployment and simultaneous execution of multiple software components on top of the same virtual machine. However, isolation between the various components is limited. A faulty version of any one of the software components can compromise the whole system by consuming all available resources. We propose a solution to efficiently identify faulty software components running simultaneously in a single virtual machine. It is based on an optimistic adaptive monitoring system to identify the faulty component. Suspected components are instrumented to obtain fined grain data for deeper analysis by the monitoring system, but only when required. Unsuspected components are left untouched and execute normally. Thus, we perform localized, just-in-time monitoring that decreases the accumulated overhead of the monitoring system. We evaluated our approach against a state-of-the-art monitoring system and we have shown that our technique correctly detects faulty components, while reducing overhead by an average of 80% [52]. Based on this work, we have presented two tutorials at the CBSE/QoSA conference [49] and at the Middleware conference [51].

### 6.4.2. Technology enablers for resource-aware dynamic software architecture

Models@runtime provides semantically rich reflection layers enabling intelligent systems to reason about themselves and their surrounding context. Most reasoning processes require not only to explore the current state, but also the past history to take sustainable decisions e.g. to avoid oscillating between states. Models@runtime and model-driven engineering in general lack native mechanisms to efficiently support the notion of history, and current approaches usually generate redundant data when versioning models, which reasoners need to navigate. Because of this limitation, models fail in providing suitable and sustainable abstractions to

deal with domains relying on history-aware reasoning. This work tackles this challenge by considering history as a native concept for modeling foundations. Integrated in conjunction with lazy load/storage techniques into the Kevoree Modeling Framework, we demonstrated onto a energy-aware smart grid case study that this mechanisms enable a sustainable reasoning about massive historized models [53].

In this field we also created a specific extension to the `docker.io` open-source project to support a dynamic resource reservation of running containers [9]

### 6.4.3. *Efficient reasoning techniques for dynamic software architecture*

Providing software with the capacity of adapting itself according to its environment requires effective techniques to reason and decide on what adaptation to undertake over the running system. To decide on a system adaptation, we have to characterize the value of the system in its corresponding execution environment. A system cannot be characterized by a single dimension, but only using several dimensions such as performance, energy consumption, security and so on. In this context, we have proposed various techniques to leverage multi-objective evolutionary algorithms both at deployment time [46], [21] and at runtime [47] to enable system optimization using multidimensional optimization. We have also proposed a technique to adapt a system proactively based on predictions in order to prevent failures [60]

### 6.4.4. *The Internet of Things application domain*

We apply our techniques for heterogeneous and dynamic software architecture more specifically to the Internet of Things application domain. We have two main contributions: (1) an application of the models@runtime concepts on embedded nodes with very limited resources for memory, CPU and battery [30], and (2) a study on the problem of renewable energy production and consumption at home [39]. Domestic microgeneration is the onsite generation of low and zero-carbon heat and electricity by private households to meet their own needs. In this paper we explore how an everyday household routine (doing laundry) can be augmented by digital technologies to help households with photovoltaic solar energy generation to make better use of self-generated energy. We present an 8 month in the field study that involved 18 UK households in longitudinal energy data collection, prototype deployment and participatory data analysis [38]. Through a series of technology interventions mixing energy feedback, proactive suggestions and direct control, the study uncovered opportunities, potential rewards and barriers for families to shift energy consuming household activities. The study highlights how digital technology can act as a mediator between household laundry routines and energy demand-shifting behaviors. Finally, the study provides insights into how a "smart" energy-aware washing machine shapes organization of domestic life and how people "communicate" with their washing machine.

## 6.5. 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 through multiple sensors, run eternally and are open to other services and to users. Here we present our latest observational and technical results about new approaches to increase diversity in software systems.

### 6.5.1. *Automatic synthesis of computationally diverse program variants*

The predictability of program execution provides attackers with a rich source of knowledge that they can exploit to spy or remotely control the program. Moving target defense addresses this issue by constantly switching between many diverse variants of a program, thus reducing the certainty that an attacker can have about the program execution. The effectiveness of this approach relies on the availability of a large number of software variants that exhibit different executions. However, current approaches rely on the natural diversity provided by off-the-shelf components, which is very limited. We have explored the automatic synthesis of large sets of program variants, called *sosies* [32]. Sosies provide the same expected functionality as the original program, while exhibiting different executions. They are said to be computationally diverse.

### 6.5.2. Software Evolution for Diversity Emergence

We aim at favoring spontaneous diversification in software systems, to increase their adaptive capacities. This objective is founded on three observations: (1) software has to constantly evolve to face unpredictable changes in its requirements, execution environment or to respond to failure (bugs, attacks, etc.); (2) the emergence and maintenance of high levels of diversity are essential to provide adaptive capacities to many forms of complex systems, ranging from ecological and biological systems to social and economical systems; (3) diversity levels tend to be very low in software systems. In this work [33], we consider evolution as a driver for diversity as a means to increase resilience in software systems. In particular, we are inspired by bipartite ecological relationships to investigate the automatic diversification of the server side of a client-server architecture.

### 6.5.3. Analyzing the diversity of development practices in open source projects

Decentralized version control systems allow a rich structure of commit histories, which presents features that are typical of complex graph models. We bring some evidences of how the very structure of these commit histories carries relevant information about the distributed development process. By means of a novel data structure that we formally define, we analyze the topological characteristics of commit graphs of a sample of git projects. Our findings point out the existence of common recurrent structural patterns that identically occur in different projects and can be considered building blocks of distributed collaborative development [36], [35].

**DREAM Project-Team**

# 6. New Results

## 6.1. Simulator-based decision support

**Participants:** Louis Bonneau de Beaufort, Tassadit Bouadi, Marie-Odile Cordier, Thomas Guyet, Christine Largouët, Véronique Masson, René Quiniou, Sophie Robin, Laurence Rozé, Yulong Zhao.

### *6.1.1. Model-checking an ecosystem for decision-aid*

In previous work we have proposed to use qualitative modelling to model ecosystems and we defined a set of high level query patterns to explore th system [53]. This approach has been applied on real-case ecosystems (coral-reef ecosystem in New-Caledonia, fisheries ecosystem in the English channel) and implemented in a tool called EcoMata.

In recent studies we have focussed on the formalization of the qualitative model automatically built from an abstracted ecosystem description. Ecosystems share some common features with concurrent systems represented in the model-checking field: the system complexity is due to interacting components and the system evolution is event-driven and submitted to temporal constraints. However if model-checking techniques are dedicated to finite state systems, ecosystem models are usually represented by analytical models as a set of differential equations. Some studies present how to quantize continuous-time systems in order to diagnose them as discrete-event systems. We proposed a method to build automatically a network of timed automata from various information on the system: description of interactions between components, human knowledge, simple models of population dynamics. The key point is to quantize the continuous-time sub-systems and to get a qualitative model described as network of timed automata. To reduce the size of this network, important after the automatic generation, a learning machine algorithm has been applied in order to reduce the number of "similar" locations. This work has been published in [37].

### *6.1.2. Controler synthesis for optimal strategy search*

Similarly to previous work, this approach relies on a qualitative model of a dynamical system. The problem consists in finding a strategy in order to help the user achieveing a specific goal. The model is now considered as a timed game automata expressing controllable and uncontrollable actions. The strategy represents the sequence of actions that can be performed by a user to reach a particular state (in case of a reachability problem for instance). A first approach based on a "generate and test" method has been developped for the marine ecosystem example [69].

More recently, two new methods for finding the optimal strategies have been proposed. The first one uses controller synthesis on timed automata and exploits the efficency of well-recognized tools. The second one deals with a set of similar models and extracts a more general strategy, closer to what is expected by the stakeholders. These methods have been applied in the context of herd management on a catchment. Yulong Zhao defended his Phd this year on this research subject [5].

### *6.1.3. A datawarehouse for simulation data*

In previous work we have proposed a datawarehouse architecture to store the huge data produced by deep agricultural simulation models [35]. This year, we have worked on hierarchical skyline queries to introduce skyline queries in a datawarehouse framework. Conventional skyline queries retrieve the skyline points in a context of dimensions with a single hierarchical level. However, in some applications with multidimensional and hierarchical data structure (e.g. data warehouses), skyline points may be associated with dimensions having multiple hierarchical levels. Thus, we have proposed an efficient approach reproducing the effect of the OLAP operators "drill-down" and "roll-up" on the computation of skyline queries [10], [25]. It provides the user with navigation operators along the dimensions hierarchies (i.e. specialize / generalize) while ensuring an online calculation of the associated skyline.

### 6.1.4. Post-mining classification rules

We consider sets of classification rules with quantitative attributes inferred by supervised machine learning, as in the framework of the SACADEAU project. Our aim is to improve human understanding of such sets of rules. Often, output quantitative rules contains too many intervals that are difficult to intepret. It is thus important to merge some of these intervals in order to get more understandable rules. However, blindly merging rules may decrease rule quality. To counter that, we proposed two algorithms for merging intervals via clustering techniques that take into account the final rule quality. The approach automatically detects the most adapted number of clusters required to merge intervals while maintaining rule quality.

## 6.2. Data Mining

**Participants:** Sid Ahmed Benabderrahmane, Marie-Odile Cordier, Serge Vladimir Emteu Tchagou, Thomas Guyet, Yves Moinard, René Quiniou, Alexandre Termier.

### 6.2.1. Application of sequential pattern mining with intervals

Our theoretical work on sequential pattern mining with intervals [47] has been applied to two real issues: the customer relationship management and analysis of care pathways.

Customer Relationship Management (CRM) comprises a set of tools for managing the interactions between a company and its customers. The main objective of the data analysts is to propose the correct service to a customer at the correct moment by applying decision rules. If rules or sequential patterns can predict the interaction that can follow a sequence of actions or events, they can not predict at what time such actions have the highest probability to occur. The objective of temporal pattern mining is to refine the prediction by extracting patterns with information about the duration and delay between the events. This year we have experimented two algorithms on a CRM databases, QTIPrefixSpan [47] and TGSP [68], to extract sequential patterns with quantitative temporal information. We have integrated the TGSP algorithm into an interface to visualize and to browse the extracted patterns. A paper describing this contribution have been recently accepted in a workshop [41].

The QTIPrefixSpan algorithm has also been applied to the analysis of care pathways. The pharmaco-epidemiology platform of the Rennes hospital was interested in characterizing the care pathways preceeding the epileptic seizures of stable epilepic patients. A care pathway consist of the sequence of drug exposures (temporal intervals). The objective is to study the ability of QTIPrefixSpan to identify drug switches between original and generic anti-epileptic drugs. This work is still in progress and will be extended in the PEPS project (see section 8.1.1 ).

### 6.2.2. Multiscale segmentation of satellite image time series

Satellite images allow the acquisition of large-scale ground vegetation. Images are available along several years with a high acquisition frequency (1 image every two weeks). Such data are called satellite image time series (SITS). In [45], we presented a method to segment an image through the characterization of the evolution of a vegetation index (NDVI) on two scales: annual and multi-year. This work is now under submission to the journal on Remote Sensing in Environment. The main issue of this approach was the required computation resources (time and memory). Last year, we applied 1D-SAX to reduce data dimensionality [21]. This approach on the supervised classification of large SITS of Senegal and we showed that 1D-SAX approaches the classification results of time series while significantly reducing the required memory storage of the images.

This year, we first continued to explore the supervised classification of SITS using classification trees for time-series [40] by implementing a parallelized version of this algorithm. Secondly, we explored the adaption of the object-oriented segmentation to SITS. The object-oriented segmentation [34] is able to segment images based on the segment uniformity. We proposed a measure for time-series uniformity to adapt the segmentation algorithm and applied it on large multivariate SITS of Senegal. This work have been presented to the conference on spatial analysis and geography [16]. A collaboration with A. Fall (Université Paris-13) have been initiated to compare our results on the Senegal with ground observations. Moreover, we planned to apply our algorithm to analyse the land use in Peru (collaboration with A. Marshall, Université Paris 13/PRODIG).

### 6.2.3. Analysis and simulation of landscape based on spatial patterns

Researchers in agro-environment need a great variety of landscapes to test their scientific hypotheses using agro-ecological models. Real landscapes are difficult to acquire and do not enable the agronomist to test all their hypothesis. Working with simulated landscapes is then an alternative to get a sufficient variety of experimental data. Our objective is to develop an original scheme to generate landscapes that reproduce realistic interface properties between parcels. This approach is made of the extraction of spatial patterns from a real geographic area and the use of these patterns to generate new "realistic" landscapes. It is based on a spatial representation of landscapes by a graph expressing the spatial relationships between the agricultural parcels (as well as the roads, the rivers, the buildings, etc.), of a specific geographic area.

In past years, we worked on the exploration of graph mining techniques, such as gSPAN [67], to discover the relevant spatial patterns present in a spatial-graph. We assume that the set of the frequent graph patterns are the characterisation of the landscape. Our remaining challenge was to simulate new realistic landscapes that will reproduce the same patterns.

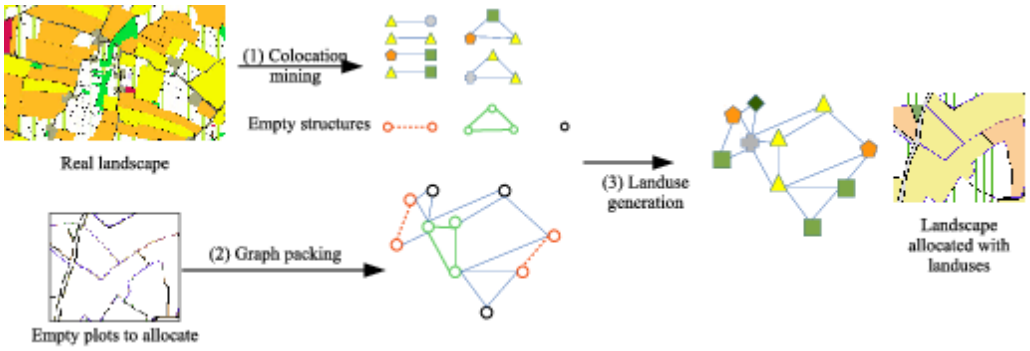


*Figure 1. Simulation process in three steps : 1) characteristic graph-patterns mining, 2) graph packing of the cadastral landscape and 3) land use assignement.*

This year, we formalized the simulation process by a formal problem of graph packing [51]. The process is illustrated by Figure 1 . Solving instances of the general graph packing problem has a high combinatorics and there does not exists any efficient algorithm to solve it. We proposed an ASP program to tackle the combinatorics of the graph packing and to assign the land use considering some expert knowledge. Our approach combines the efficiency of ASP to solve the packing issue and the simplicity of the declarative programming to take into account the expert contraints on the land use. Contraints about the minimum surface of crops or about the impossibility of some crops colocation can be easily defined. This work have been presented at the conference RFIA [19] and we have been invited to provide an extended version to the Revue d'Intelligence Artificielle (RIA). The application results have been presented to the national colloquium on landscape modelling (http://www.reseau-payote.fr/?q=colloque2014).

In addition to the landscape simulation, the challenging tasks of solving the general graph packing with ASP raises interests in more general problem (such as graph compression). We initiated a collaboration with J. Nicolas (Inria/Dyliss) to improve the efficiency of our first programs.

### 6.2.4. Mining with ASP

In pattern mining, a pattern is considered interesting if it occurs frequently in the data, i.e. the number of its occurrences is greater than a fixed given threshold. As non informed mining methods tend to generate massive results, there is more and more interest in pattern mining algorithms able to mine data considering some expert knowledge. Though a generic pattern mining tool that could be tailored to the specific task of a data-scientist

is still a holy grail for pattern mining software designers, some recent attempts have proposed generic pattern mining tools [44] for itemset mining tasks. In collaboration with Torsten Schaub, we explore the ability of a declarative language, such as Answer Set Programming (ASP), to solve pattern mining tasks efficiently. A first attempt have been proposed by Jarvisälo for simple settings [49].

This year, we worked on several classical pattern mining tasks: episodes, sequences and closed/maximal itemsets. In [20], we explore the use of ASP to extract frequent episodes (without parallel events) in a unique long sequence of itemsets. We especially evaluate the incremental resolution to improve the efficiency of our program. We next worked on sequence mining to extract pattern from the sequence of TV programs (V. Claveau, CNRS/LinkMedia). This tasks was simpler, but the computation time was significantly higher than dedicated algorithms. Nonetheless, our recent programs extracting closed or maximal patterns have better results.

### 6.2.5. *Monitoring cattle*

Following the lines of a previous work [62], we are working on a method for detecting Bovine Respiratory Diseases (BRD) from behavioral (walking, lying, feeding and drinking activity) and physiological (rumen temperature) data recorded on feedlot cattle being fattened up in big farms in Alberta (Canada). This year, we have especially worked on multivariate sensor analysis to devising multivariate decision rules for improving the specificity of detectors [15].

### 6.2.6. *Subdimensional clustering for fast similarity search over time series data. Application to Information retrieval tasks*

Information retrieval and similarity search tasks in time series databases remains a challenge that requires to discover relevant pattern-sequences that are recurrent over the overall time series sequences, and to find temporal associations among these frequently occurring patterns. However, proposed methods suffer from a lack of flexibility of the used similarity measures, a lack of scalability of the representation model, and a penalizing runtime to retrieve the information. Motivated by these observations, we have designed a framework tackling the query by content problem on time series data, ensuring (i) fast response time, (ii) multi-level information representation, and (iii) representing temporal associations between extracted patterns. This year we have compared several distance measures on time series with different criteria and proposed a hybrid retrieval method based on pattern extraction and clustering [8].

### 6.2.7. *Knowledge Extraction from Heterogeneous Data*

Recently, mining microarrays data has become a big challenge due to the growing sources of available data. We are using machine learning methods such as clustering, dimensionality reduction, association rules discovery on transcriptomic data, by combining a domain ontology as source of knowledge, in order to supervise the KDD process. Our objectives concern the identification of genes that could participate in the development of tumors. This year, we have introduced a new method for extracting enriched biological functions from transcriptomic databases using an integrative bi-classication approach based on formal concept analysis [7].

### 6.2.8. *Trace reduction*

One problem of execution trace of applications on embedded systems is that they can grow very large, typically several Gigabytes for 5 minutes of audio/video playback. Some endurance tests require continuous playback for 96 hours, which would lead to hundreds of Gigabytes of traces, that current techniques cannot analyze. We have proposed TraceSquiz, an online approach to monitor the trace output during endurance test, in order to record only suspicious portions of the trace and discard regular ones. This approach is based on anomaly detection techniques, and as been accepted in the DATE'15 conference [14].

## 6.3. Causal reasoning and argumentation

**Participants:** Philippe Besnard, Marie-Odile Cordier, Yves Moinard.

We have continued our work on reasoning (precisely search for explanations) from causal relations and ontology [9]. Mainly, we have enforced the use or argumentation [32] in order to help choosing the best explanations among the (rather big) set of explanations given by our previous formalism. Then, we hope to be able to use the last versions of clingo in order to get an efficient tool to deal with complex situations (our example is Xynthia storm, february 2012 in western France for which there exists a huge amount of data from various official reports) by using clingo. For now we have a preliminary program which provides (besides the applications already mentioned: mining and landscape simulation) another application of the recent versions of ASP. One interest is that the nature of the rules in ASP should allow to translate rather directly (and hopefully efficiently) our previous formalism together with the improved argumentation part.

**DYLISS Project-Team**

# 6. New Results

## 6.1. Highlights of the Year

Four PhD theses were defended this year. They evidenced that ASP-technologies are now mature enough to perform data integration of large-scale bio-molecular datasets: classification of families of proteins [10], reconstruction of regulatory networks [13], reconstruction of metabolic network [11], and modelling of the discrete dynamics of a signalling or a regulatory network [12]. Importantly, symbolic classification technics have been adapted to exhibit relevant biological features: we used both formal concept analysis and semantic-based analysis for sequence and network analysis.

## 6.2. Data integration

**Participants:** Jacques Nicolas, Charles Bettembourg, Jérémie Bourdon, Jeanne Cambefort, Marie Chevallier, Guillaume Collet, Olivier Dameron, Damien Eveillard, Julie Laniau, Sylvain Prigent, Anne Siegel, Valentin Wucher.

**Pan-genomic metabolic network of *Ectocarpus siliculosus***: We introduced the first metabolic network for the non-classical species *E. Siliculosus*, called EctoGEM. The reconstruction process includes draft reconstruction based on sequence and functional annotation analysis. It is followed by a combinatorial gap-filling process using the Meneco software based on answer set programming, a semantic analysis of the completion and a manual curation. This reconstruction enables a better understanding of organism biology and a reannotation of its genome. [*J. Cambefort, G. Collet, O. Dameron, D. Eveillard, S. Prigent, A. Siegel*] [22], [11]

**New insights on bacteria associated with brown algae** As an application of our tools for the reconstruction of metabolic networks, we have contributed to the analysis of the genome of a bacteria which lives in symbiosis with brown algae by investigating candidates for metabolic exchanges between the bacteria and the algae. [*G. Collet, J. Cambefort, A. Siegel*] [19] [Online publication]

**Modeling parsimonious putative regulatory networks** We integrated heterogeneous information from two types of network predictions to determine a causal explanation for the observed gene co-expression. We modeled this integration as a combinatorial optimization problem. We demonstrated that this problem belongs to the NP-hard complexity class. We proposed an heuristic approach to have an approximate solution in a practical execution time. Our evaluation showed that the E.coli regulatory network resulting from the application of this method has higher accuracy than the putative one built with traditional tools. [*A. Aravena, A. Siegel, D. Eveillard*] [26] [Online publication]

**Modeling of a gene network between mRNAs and miRNAs to predict gene functions involved in phenotypic plasticity in the pea aphid and non coding RNA in pea aphid** During its PhD, V. Wucher has built the first network model of gene regulation by microRNAs in pea aphid. The thesis has studied the discrimination between embryos development towards either sexual or asexual reproduction types in the pea aphid *Acyrthosiphon pisum*, at the genomic level. The study of the post-transcriptional regulation network implies both the identification of regulated elements during embryogenesis and the identification of the interaction modules between microRNAs and mRNAs using formal concept analysis. It helps the understanding of regulation differences between sexual and asexual embryogenesis. Moreover, it is the first step towards the modeling of the entire set of genes regulations in pea aphid during embryogenesis. [*V. Wucher, J. Nicolas, F. Legeai (Genscale team)*] [13], [39], [30]

**Using a large-scale knowledge database on reactions and regulations to exhibit key regulators** A new formalism of regulated reactions combining biochemical transformations and regulatory effects was proposed to unify the different mechanisms contained in knowledge libraries. Based on a related causality graph, an algorithm was developed to propose a reasonable set of upstream regulators from lists of target molecules. Scores were added to candidates according to their ability to explain the greatest number of targets or only few specific ones. The method was validated on a real example related to glycolysis. [*P. Blavy, A. Siegel*] [18] [Online publication]

**Semantic particularity measure for functional characterization of gene sets using gene ontology** We propose a new approach to compute gene set particularities based on the information conveyed by Gene Ontology terms. A GO term informativeness can be computed using either its information content based on the term frequency in a corpus, or a function of the term's distance to the root. We demonstrated that the combination of semantic similarity and semantic particularity measures was able to identify genes with particular functions from among similar genes. This differentiation was not recognized using only a semantic similarity measure. [*C. Bettembourg, O. Dameron*] [17] [Online publication]

**Integrating GALAXY workflows in a metadata management environment** New tools are needed to enable the quick design and the intensive parallel execution of bioinformatics processes. Therefore, we proposed a new dataflow-oriented workflow management system dedicated to intensive bioinformatics tasks. We worked on the interoperatibility of bioinformatics workflows using a model-driven approach. Our results enable new import / export capabilities beetwen multiple workflow management environnements and insights to create a unique shared workflow model. [*O. Dameron, F. Moreews (Genscale team), Y. Le Bras (GenOuest platform), C. Monjeaud (GenOuest platform), O. Collin (GenOuest platform)*][36]

## 6.3. Time-series and asymptotic dynamics

**Participants:** Anne Siegel, Jérémie Bourdon, Jeanne Cambefort, Damien Eveillard, Vincent Picard, Nathalie Théret, Santiago Videla.

**Reasoning on the response of logical signaling networks with boolean models** A series of papers and a PhD thesis focused on modeling the response of logical signaling networks by means of automated reasoning using ASP. In this context, a crucial issue is automatic learning of logical networks from partial observations of input/output behaviours, in order to achieve unbiased and robust discoveries. Experiments showed that many networks can be compatible with a given set of experimental observations. In a review chapter, we first discuss how ASP can be used to exhaustively enumerate all these logical networks. Next, in order to gain control over the system, we look for intervention strategies that force a set of target species into a desired steady state. Finally, we discuss the usage of ASP for solving the aforementioned problems and the novelty of our approach with respect to existing methods. [*S. Videla, A. Siegel, J. Nicolas*] [23], [38], [12] [Online publication]

**Integrative modeling framework for signaling networks based on guarded transitions models** We develop a new non-ambiguous formal interpretation of signaling pathways as discrete dynamic models. The resulting language, Computer-Aided Design for BIOlogical Models (CADBIOM), is based on a simplified version of guarded transitions in which we introduced temporal parameters for each transition to manage competition and cooperation between parts of the models . Tools for simulation and model checking analyses using the formal Cadbiom language have been developed (http://cadbiom.genouest.org). Using CADBIOM, we built the first discrete model of TGF-$\beta$ signaling networks by automatically integrating the 137 human signaling maps from the Pathway Interaction Database into a single unified dynamic model. Temporal property-checking analyses of 15934 trajectories that regulate 145 TGF-$beta$ target genes reveal the association of specific pathways with distinct biological processes. [*G Andrieux, M Le Borgne, N. Théret*] [15] [Online publication]

**Exploring metabolism flexibility in complex organisms through quantitative study of precursor sets for system outputs** We extended a Flux-Balanced-Analysis approach to quantify the precursor composition of each system output and to discuss the biological relevance of a set of flux in a given metabolic network. The composition is called contribution of inputs over outputs [AIO]. In order to further investigate metabolic network flexibility, we have proposed an efficient local search algorithm computing the extremal values of

AIO coefficients. This approach enables to discriminate diets without making any assumption on the internal behaviour of the system. [*O. Abdou-Arbi, J. Bourdon, A. Siegel*] [14] [Online publication]

**Multivariate Normal Approximation for the Stochastic Simulation Algorithm: limit theorem and applications** We prove a central limit theorem on the asymptotic stochastic dynamical behavior of the outputs of a reaction network under certain steady-state assumptions. We present multiple applications including a constraints-based approach to verify asymptotic properties on the output moments without prior knowledge about the kinetic parameters. [*V. Picard, A. Siegel, J. Bourdon*] [33] [Online publication]

**Numeric model for initiation of translation in sea-urchin** We use a numeric-based modeling approach to study the regulation of protein synthesis following fertilization in sea urchin. This approach based on parcimonious modelling evidenced that two processes are required to explain experimental data: a destabilization of eIF4E:4E-BP complex and a great stimulation of the 4E-BP-degradation mechanism, both rapamycin-sensitive [*A. Siegel, J. Bourdon*] [21] [Online publication]

## 6.4. Sequence annotation

**Participants:** François Coste, Aymeric Antoine-Lorquin, Catherine Belleannée, Guillaume Collet, Gaëlle Garet, Clovis Galiez, Laurent Miclet, Jacques Nicolas, Valentin Wucher.

**Automated Enzyme Classification by Formal Concept Analysis** Guessing enzyme's functional activity from its sequence is a crucial task that can be approached by comparing the new sequences with those of already known enzymes labeled by a family class. This task is difficult because the activity is based on a combination of small sequence patterns and sequences greatly evolved over time. We have designed a classifier based on the identification of common subsequence blocks between known and new enzymes and the search of formal concepts built on the cross product of blocks and sequences for each class. Since new enzyme families may emerge, it is important to propose simultaneously a first classification of enzymes that cannot be assigned to a known family. Formal Concept Analysis offers a nice framework to set this task as an optimization problem on the set of concepts. The classifier has been tested with success on a particular set of enzymes present in a large variety of species, the haloacid dehalogenase (HAD) superfamily. [*F. Coste, G. Garet, J. Nicolas*] [28], [10]

**A bottom-up efficient algorithm learning substitutable languages from positive examples** Based on Harris's substitutability criterion, the recent definitions of classes of substitutable languages have led to interesting polynomial learnability results for expressive formal languages. These classes are also promising for practical applications: in natural language analysis, because definitions have strong linguisitic support, but also in biology for modeling protein families, as suggested in our previous study introducing the class of local substitutable languages. But turning recent theoretical advances into practice badly needs truly operable algorithms. We present here an efficient learning algorithm, motivated by intelligibility and parsing efficiency of the result, which directly reduces the positive sample into a small non redundant canonical grammar of the target substitutable language. Thanks to this new algorithm, we have been able to extend our experimentation to a complete protein dataset confirming that it is possible to learn grammars on proteins with high specificity and good sensitivity by a generalization based on local substitutability. [*F. Coste, G. Garet, J. Nicolas*] [29], [10]

**Logol: Expressive Pattern Matching in sequences. Application to Ribosomal Frameshift Modeling** Logol consists in both a language for describing biological patterns, and an associated parser for effective pattern search in sequences (RNA, DNA or protein). The Logol language, based on an high level grammatical formalism (String Variable Grammars), allows to express flexible patterns (with mispairings and indels) composed of both sequential elements (such as motifs) and structural elements (such as repeats or pseudoknots). Its expressive power allows the design of sophisticated patterns such as the signature of "-1 programmed ribosomal frameshifting" (PRF) events in messenger RNA sequences. A PRF signature is a complex model composed of a slippery site followed by a pseudoknot located in a specific part of the sequence, which provides a good illustration of the Logol language power. [*C. Belleannée, J. Nicolas, O. Sallou (GenOuest platform)*] [27] [Online publication]

**Identifying distant homologous viral sequences in metagenomes using protein structure information** It is estimated that marine viruses daily kill about 20% of the ocean biomass. Identifying them in water samples is thus a biological issue of great importance. The metagenomic approach for virus identification is a challenging task since their sequences carry a lot of mutations and are very difficuly to identify by standard homology searches. The PEPS VAG project aims at establishing a novel methodology that uses structures of proteins as extra-information in order to annotate metagenomes without relying on sequence homology. In the context of the first experiments made on the metagenome of station 23 of the TARA Ocean Project, we used the structures of capsid proteins to infer the sequence signature of their fold, in order to find them in the metagenome. This work presents the methodology, the first experiments and the on-going improvements. [*C. Galiez, F. Coste*] [35]

**Computational Protein Design: trying an Answer Set Programming approach to solve the problem** The problem of *Computational Protein Design* aims at finding the best protein conformation to perform a given task. This problem can be reduced to an optimization problem, looking for the minimum of an energy function depending on the amino-acid interactions in the protein. The CPD problem may be easily modeled as an ASP program but a practical implementation able to work on real-sized instances has never been published. We have raised the main source of difficulty for current ASP solvers and ran a series of benchmarks highlighting the importance of finding a good upper bound estimation of the target minimum energy to reduce the amount of combinatorial search. Our solution clearly outperforms a direct ASP implementation without this estimation and has comparable performances with respect to SAT-based approaches. It remains less efficient than a recent approach by cost function networks, showing there still exists some place for improving the optimization component in ASP with more dynamical strategies. [*J. Nicolas, H. Bazille*] [34]

**Searching for Optimal Orders for Discretized Distance Geometry** The Molecular Distance Geometry Problem (MDGP) is the problem of finding the possible conformations of a molecule by exploiting available information on some distances between pairs of its atoms. When some assumptions are satisfied, the MDGP can be discretized, so that the search domain of the problem becomes a tree where each node corresponds to a candidate position for an atom. The search tree can be efficiently explored by using an *interval* Branch & Prune ($i$BP) algorithm that can potentially enumerate all feasible conformations. In this context, the order given to the atoms of the molecule plays an important role, because it allows the discretization assumptions to be satisfied, and it also impacts the computational cost of the $i$BP algorithm. We have proposed a new discretization order for protein backbones based on the optimization of certain criteria for a faster exploration of the discretized search domain. To this aim, we express the search for optimal orders by a set of logical constraints in ASP. Our comparison with previously proposed orders for protein backbones shows that this new discretization order makes $i$BP perform better. [*J. Nicolas, A. Muccherino (Genscale Team)*] [43]

**From analogical proportions in lattices to proportional analogies in formal concepts** We provided an attempt at bridging formal concept analysis and the modeling of analogical proportions (i.e., statements of the form “a is to b as c is to d”). A suitable definition for analogical proportions in non distributive lattices is proposed and then applied to concept lattices. This enables us to compute what we call proportional analogies. In addition, we define the locally maximal subwords and locally minimal superwords common to a finite set of words. We also define the corresponding sets of alignments. We show that the constructed family of sets of alignments has the lattice structure. The study of analogical proportion in lattices gives hints to use this structure as a machine learning basis, aiming at inducing a generalization of the set of words. [*L. Miclet*] [32], [37]

**ESTASYS Exploratory Action**

# 6. New Results

## 6.1. Highlights of the Year

The Plasma statistical model checker has been made available to other scientists. ESTASYS has open a new branch on verifying the security of complex systems.

## 6.2. Verification of Heterogeneous Systems

**Participants:** Axel Legay, Benoît Boyer, Ngo Van-Chan, Jean Quilbeuf.

> This part concerns Tasks 1, 2 and 4 of the action. We characterize and formalize heterogeneous aspects of SoS and then we define efficient monitoring algorithms and representations for their requirements. We then combine the results with Statistical Model Checking (Task 5).

Systems of Systems (SoS) are very large scale systems with particular characteristics. SoS are not directly built from scratch by a single designer or a single team but are obtained as the composition of simpler systems. SoS have strong reliability and dependability requirements, as they aim to provide a service over a long running period. SoS may dynamically modify themselves by connecting to new systems, updating or disconnecting faulty ones, making it impossible to statically know the set of subsystems that are part of the SoS before runtime.

One of the main difficulty arising when developing SoS is the fact that subsystems may have been designed with a different goal in mind. In particular, some subsystems may have their own goal which differs from the global goal of the SoS. Furthermore, each subsystem may be developed in a particular computation model, making it difficult to find a common unifying semantics for the whole SoS. Finally, SoS may exhibit some emergent behaviors that are hardly predictable at design time.

One of the solutions to allow simulation of a SoS is to rely on a common interface for interconnecting the subsystems. The Functional Mockup Interface (FMI) standard is a natural candidate for such an interface. The different components of a SoS developed in different models of computation can be translated to Functional Mockup Units (FMU). Then a so-called master algorithm coordinates the FMUs composing the system. The execution of each FMU is either directly handled by the master algorithm or relies on an external tool for its execution.

Because the subsystems composing a SoS are of heterogeneous nature, it is difficult to find a common semantics model for the whole system. Furthermore, building such a transition system is not tractable due to the complexity of the system. Thus verification through traditional model checking is not possible for SoS. However, since the FMI/FMU framework enables simulation of such systems, the statistical model checking approach can be used.

The DANSE EU project aims to provide a complete tool chain from the modeling to the verification of SoS. At the higher level, the modeling is done in UPDM using the RHAPSODY tool. At the same level, the designer can express requirements over the model using some patterns written in GCSL. The UPDM model can then be translated into a FMI/FMU format that can be simulated by a dedicated tool, named DESYRE. Similarly, the GCSL requirements are transformed into BLTL formulas. Finally, the PLASMA statistical model checker has been integrated with the DESYRE tool chain in order to check the BLTL formulas based on the simulations provided by DESYRE.

### *6.2.1. Papers:*

[45] (W) This report presents some of the results of the first year of Danse, one of the first EU IP projects dedicated to System of Systems. Concretely, we offer a tool chain that allows to specify SoS and SoS requirements at high level, and analyse them using powerful toolsets coming from the formal verification area. At the high level, we use UPDM, the system model provided by the british army as well as a new type of contract based on behavioral patterns. At low level, we rely on a powerful simulation toolset combined with recent advances from the area of statistical model checking. The approach has been applied to a case study developed at EADS Innovation Works.

[51] (W) Exhaustive formal verification for systems of systems (SoS) is impractical and cannot be applied on a large scale. In this paper we propose to use statistical model checking for efficient verification of SoS. We address three relevant aspects for systems of systems: 1) the model of the SoS, which includes stochastic aspects; 2) the formalization of the SoS requirements in the form of contracts; 3) the tool-chain to support statistical model checking for SoS. We adapt the SMC technique for application to heterogeneous SoS. We extend the UPDM/SysML specification language to express the SoS requirements that the implemented strategies over the SoS must satisfy. The requirements are specified with a new contract language specifically designed for SoS, targeting a high-level English-pattern language, but relying on an accurate semantics given by the standard temporal logics. The contracts are verified against the UPDM/SysML specification using the Statistical Model Checker (SMC) PLASMA combined with the simulation engine DESYRE, which integrates heterogeneous behavioral models through the functional mock-up interface (FMI) standard. The tool-chain allows computing an estimation of the satisfiability of the contracts by the SoS. The results help the system architect to trade-off different solutions to guide the evolution of the SoS.

## 6.3. Formal Models for Variability

**Participants:** Axel Legay, Rudolf Fahrenberg, Jin Hyun Kim.

> This part of the report is more concerned with task 2. It studies variability aspects in the broad scope. To simplify the study for the first year, we use the concept of software product lines. Later we shall use the results in federation of embedded systems, which is a particular class of Systems of systems.

Variability is ubiquitous in today's systems, be it in the form of configuration options or extensible architectures. By mastering variability, developers can adapt their system to changing requirements without having to develop entirely new applications. Variability is central in the context of SoS whose behaviors depend on interconnected objects. To gain information on managing variability, we have focused on Software Product Lines. Software Product Lines (SPLs) are a popular form of variability-intensive systems. They are families of similar software systems developed together to make economies of scale. SoS can be viewed as examples of product lines with interconnected obbjects. SPL engineering aims to facilitate the development of the members of a family (called *products* or *variants*) by identifying upfront their commonalities and differences. Variability in SPLs is commonly represented in terms of *features*, *i.e.*, units of difference between products that appear natural to stakeholders. Each product of an SPL is therefore defined by its set of features. Hierarchies of features and dependencies between features (*e.g.*, requires, excludes) are typically captured in a *Feature Model* (FM), *i.e.* a tree-like structure that specifies which combinations of features are valid.

### *6.3.1. Papers:*

[15] (C) The model-checking problem for Software Products Lines (SPLs) is harder than for single systems: variability constitutes a new source of complexity that exacerbates the state-explosion problem. Abstraction techniques have successfully alleviated state explosion in single-system models. However, they need to be adapted to SPLs, to take into account the set of variants that produce a counterexample. In this paper, we apply CEGAR (Counterexample-Guided Abstraction Refinement) and we design new forms of abstraction specifically for SPLs. We carry out experiments to

evaluate the efficiency of our new abstractions. The results show that our abstractions, combined with an appropriate refinement strategy, hold the potential to achieve large reductions in verification time, although they sometimes perform worse. We discuss in which cases a given abstraction should be used.

- 

[18] (C) In this work, We explore how ideas of statistical testing, based on a usage model (a Markov chain), can be used to extract configurations of interest according to the likelihood of their executions. These executions are gathered in featured transition systems, compact representation of SPL behaviour. We discuss possible scenarios and give a prioritization procedure validated on a web-based learning management software.

## 6.4. Statistical Model Checking

**Participants:** Axel Legay, Sean Sedwards, Benoît Boyer, Louis-Marie Traonouez, Kevin Corre.

This section covers Tasks 4 and 5 of the action. It consists in developping Simulation based techniques and efficient statistical algorithms for SoS.

The use of test cases remains the default means of ensuring the correct behaviour of systems in industry, but this technique is limited by the need to hypothesise scenarios that cause interesting behaviour and the fact that a reasonable set of test cases is unlikely to cover all possible eventualities. Static analysis is more thorough and has been successful in debugging very large systems, but its ability to analyse complex dynamical properties is limited. In contrast, model checking is an exhaustive technique that verifies whether a system satisfies a dynamical temporal logic property under all possible scenarios. For nondeterministic and probabilistic systems, numerical model checking quantifies the probability that a system satisfies a property. It can also be used to quantify the expected cost or reward of sets of executions.

Numerical model checking gives precise, accurate and certain results by exhaustively exploring the state space of the model, however the exponential growth of the state space with system size (the ‘state explosion problem) typically limits its applicability to “toy” systems. Symbolic model checking using efficient data structures can make certain very large models tractable. It may also be possible to construct simpler but behaviourally equivalent models using various symmetry reduction techniques, such as partial order reduction, bisimulation and lumping. If a new system is being constructed, it may be possible to guarantee the overall behaviour by verifying the behaviour of its subcomponents and limiting the way they interact. Despite these techniques, however, the size, unpredictability and heterogeneity of real systems usually make numerical techniques infeasible. Moreover, even if a system has been specified not to misbehave, it is nevertheless necessary to check that it meets its specification.

Simulation-based approaches are becoming increasingly tractable due to the availability of high performance parallel hardware and algorithms. In particular, statistical model checking (SMC) combines the simplicity of testing with the formality of numerical model checking. The core idea of SMC is to create multiple independent execution traces of a system and count how many satisfy a property specified in temporal logic. The proportion of satisfying traces is an estimate of the probability that the system satisfies the property. By thus modelling the executions of a system as a Bernoulli random variable, the absolute error of the estimate can be bounded using, for example, a confidence interval or a Chernoff bound. It is also possible to use efficient sequential hypothesis testing, to decide with specified statistical confidence whether the probability of a property is above or below a given threshold. Since SMC requires multiple independent simulations, it may be efficiently divided on parallel computer architectures, such as grids, clusters, clouds and general purpose computing on graphics processors (GPGPU).

Knowing a result with less than 100% confidence is often sufficient in real applications, since the confidence bounds may be made arbitrarily tight. Moreover, a swiftly achieved approximation may prevent a lot of wasted time during model design. For many complex systems, SMC offers the only feasible means of quantifying performance. Historically relevant SMC tools include APMC, YMER and VESTA. Well-established numerical model checkers, such as PRISM and UPPAAL, are now also including SMC engines. Dedicated SMC tools

under active development include COSMOS and our own tool PLASMA. Recognising that SMC may be applied to any discrete event trace obtained by stochastic simulation, we have devised PLASMA-lab, a modular library of SMC algorithms that may be used to construct domain-specific SMC tools. PLASMA-lab has become the main vehicle of our ongoing development of SMC algorithms.

Our group is devising cutting edge techniques for SMC. In particular, we are developing new learning algorithms (Sect. 6.4.3 ), algorithms for nondeterministic systems (Sect. 6.4.1 ), and algorithms for rare events (Sect. 6.4.2 ).

### *6.4.1. Algorithms for Nondeterminism*

Nondeterministic models are of fundamental importance in defining complexity and are useful models of concurrency optimisation problems. This latter application is of particular importance in the context of systems constructed from subsystems ("Systems of Systems") that interact in an unpredictable way. Verifying or optimising such systems is problematic for numerical techniques because the state space is typically intractable. Nondeterminism is challenging for simulation-based techniques because, by definition, an executable semantics is not determined.

We have thus begun a line of research to develop SMC algorithms for nondeterministic systems. Our initial focus is Markov decision processes (MDP), however we are in the process of extending our work to various nondeterministc timed automata. Recent attempts to provide approximative algorithms for MDPs either do not address the standard verification problems, consider only a "spurious" subset of the standard problems or contain significant misconceptions and limitations.

In [28], we presented the first complete set of scalable SMC algorithms for MDPs. Our techniques are based on the idea of encoding a history-dependent scheduler as the seed of a pseudo-randomised hash function. Schedulers are thus chosen at random by selecting random seeds. The possibly infinite behaviour of the scheduler is completely encoded in $\mathcal{O}(1)$ memory. We presented simple sampling algorithms to find optimal schedulers and constructed the statistical confidence bounds necessary to find the optima of multiple estimates.

In [34] we devised the notion of "smart sampling" to dramatically improve the performance of the simple algorithms presented in [28]. The basic idea is to use part of the simulation budget to generate a crude estimate of the optimal scheduler and to use this information to better allocate the remaining budget. We successfully applied our algorithms to a number of standard case studies from the literature. We also highlighted the limitations of our approach.

The algorithms in [28], [34] find schedulers that minimise or maximise the probability of a property. In [37] we have adapted our algorithms to minimise or maximise the expected reward of a system. This adaptation is not entirely straightforward because the standard definition of reward properties assumes an exhaustive exploration of the state space of the MDP. We have included an implicit hypothesis test to include this assumption. In other respects optimising rewards is less challenging than optimising probabilities because rewards are effectively based on properties having probability 1. We demonstrate the accuracy of our rewards-based algorithms on standard case studies from the literature.

### *6.4.2. Rare Events in SMC*

Rare properties are often highly relevant to system performance (e.g., bugs and system failure are required to be rare) but pose a problem for statistical model checking because they are difficult to observe. Fortunately, rare event techniques such as *importance sampling* and *importance splitting* may be successfully applied to statistical model checking.

In a previous work [50], we explicitly considered the use of importance sampling in the context of statistical model checking. We presented a simple algorithm that uses the notion of cross-entropy to find the optimal parameters for an importance sampling distribution. In contrast to previous work, our algorithm uses a low dimensional vector of parameters to define this distribution and thus avoids the often intractable explicit representation of a transition matrix. We showed that our parametrisation leads to a unique optimum and can produce many orders of magnitude improvement in simulation efficiency. We demonstrated the efficacy of our methodology by applying it to models from reliability engineering and biochemistry.

Our contribution [49] was the first attempt to use importance splitting with SMC to overcome the Rare Event problem. The basic idea is to decompose a logical property into nested properties whose probabilities are easier to estimate. Importance splitting achieves this by estimating a sequence of conditional probabilities, whose product is the required result. To apply this idea to model checking it is necessary to define a score function based on logical properties, and a set of levels that delimit the conditional probabilities. We described the necessary and desirable properties of score functions and levels. We illustrated how a score function may be derived from a property and gave two importance splitting algorithms: one that uses fixed levels and one that discovers optimal levels adaptively.

### *6.4.3. SMC with Changes and Simulink*

We have proposed a new SMC algorithm for detecting probability changes in dynamic systems. We have adapted CUSUM, an algorithm that can be used to detect changes in signal monitoring. We show that CUSUM can be used to detect when the probability to satisfy a given property drops below some value. This algorithm offers new possibilities to detect, e.g., emergent behaviors in dynamic systems. Our main contributions has been to extend temporal logic with a change-based operator.

All these SMC algorithms are implemented in PLASMA-Lab, and have been recently exported to MATLAB/Simulink – a widely used environment for modeling, simulating and analyzing multidomain dynamic systems – through an integration of MATLAB/Simulink and PLASMA-lab. This integration exploit MATLAB Control, a library allowing to interact with MATLAB from Java. We have developed two different methods to link the two environments. The first method includes a new plugin for PLASMA-lab that allows to load and execute Simulink models within PLASMA-lab, and therefore apply SMC algorithms to these models. The second method proposes an application that can be launched directly within MATLAB and provide the PLASMA-Lab SMC algorithms.

We have submitted a paper [41] that presents the new CUSUM algorithm and the integration between PLASMA-Lab and Simulink. In this paper, we apply these results to a case-study developed with Simulink that models a temperature controller of a pig shed. We show how to use PLASMA-Lab to check SMC requirements, perform parameters optimisation and detect failures in the model using the new CUSUM algorithm.

### *6.4.4. Papers*

[48] (C) Statistical model checking (SMC) offers the potential to decide and quantify dynamical properties of models with intractably large state space, opening up the possibility to verify the performance of complex real-world systems. Rare properties and long simulations pose a challenge to this approach, so here we present a fast and compact statistical model checking platform, PLASMA, that incorporates an efficient simulation engine and uses importance sampling to reduce the number and length of simulations when properties are rare. For increased flexibility and ef-ficiency PLASMA compiles both model and property into bytecode that is executed on an in-built memory-efficient virtual machine.

[47] (C) We present PLASMA-lab, a statistical model checking (SMC) library that provides the functionality to create custom statistical model checkers based on arbitrary discrete event modelling languages. PLASMA-lab is written in Java for maximum cross-platform compatibility and has already been incorporated in various performance-critical software and embedded hardware platforms. Users need only implement a few simple methods in a simulator class to take advantage of our efficient SMC algorithms. PLASMA-lab may be instantiated from the command line or from within other software. We have constructed a graphical user interface (GUI) that exposes the functionality of PLASMA-lab and facilitates its use as a standalone application with multiple 'drop-in' modelling languages. The GUI adds the notion of projects and experiments, and implements a simple, practical means of distributing simulations using remote clients.

[41] (C; submitted) Statistical Model Checking (SMC) is a powerful and widely used approach that consists in extracting global information on the system by monitoring some of its executions. In this paper, we add two new stones to the cathedral of results on SMC, that are 1. a new algorithm to detect emergent behaviors at runtime, and 2. an integration of Plasma Lab, a powerful SMC checker,

as a library of Simulink. Our results are illustrated on a realistic case study.

[26] (C) In this paper, we make use of the notion of a *score function* to improve the granularity of a logical property. We show that such a score function may take advantage of heuristics, so long as it also rigorously respects certain properties. To demonstrate our importance splitting approach we present an optimal adaptive importance splitting algorithm and an heuristic score function. We give experimental results that demonstrate a significant improvement in performance over alternative approaches.

[43] (C; submitted) We introduce feedback-control statistical system checking (FC-SSC), a new approach to statistical model checking that exploits principles of feedback-control for the analysis of cyber-physical systems (CPS). FC-SSC uses stochastic system identification to learn a CPS model, importance sampling to estimate the CPS state, and importance splitting to control the CPS so that the probability that the CPS satisfies a given property can be efficiently inferred. We illustrate the utility of FC-SSC on two example applications, each of which is simple enough to be easily understood, yet complex enough to exhibit all of FC-SCC's features. To the best of our knowledge, FC-SSC is the first statistical system checker to efficiently estimate the probability of rare events in realistic CPS applications or in any complex probabilistic program whose model is either not available, or is infeasible to derive through static-analysis techniques.

## 6.5. Quantitative Reasoning

**Participants:** Axel Legay, Rudolf Fahrenberg, Louis-Marie Traonouez.

This part is concerned with Tasks 1 and 2. Mostly, we focus on quantifying properties of interconnected objects such as CPS (SoS and CPS share a lot of commonalities).

Model checking of systems deals with the question whether a given model of a computer system satisfies the properties one might want to require of it. This is a well-established and successful approach to formal verification of safety-critical computer systems.

When the models of the systems contain quantitative information, which is needed to represent the material on which the SoS is running, the model checking problem becomes complicated by the fact that in most cases, quantitative properties of the systems do not need to be satisfied exactly. Indeed, the model or the properties might be subject to measurement error, or probabilistic information might only be an approximation. In this case, it is of little use to know whether or not a model satisfies a specification precisely; what is needed instead is a notion of *satisfaction distance*: a measure which can assess to which extent a quantitative model satisfies a quantitative specification.

In other words, what is needed is a notion of satisfaction which is robust in the sense that small deviations in the model or the specification only lead to small changes in the outcome of the model checking question.

For reasoning about distributed systems or **systems-of-systems**, an important role is played by specification theories. Such systems are often far too complex to reason about, or model-check, as a whole, and additionally they might be composed of a large number of components which are implemented by different vendors. Hence one needs methods for compositional reasoning, which allow to infer properties of a system from properties of its components, and for incremental design, which allow to synthesize and refine specifications in a step-wise manner.

Such specification theories are by now well-established e.g. in the incarnations of interface theories and (disjunctive) modal transition systems. Additionally to defining a formalism for describing and model-checking specifications, they provide notions of refinement of specifications, logical conjunction of specifications, and structural composition and quotient.

When the models and specifications contain quantitative information, all the above notions need to be made robust. One needs to introduce a quantitative version of refinement, and the operations on specifications need to be continuous with respect to refinement distance: compositions of specifications with small refinement distance need themselves to have small refinement distance.

### *6.5.1. Theory papers:*

[33] (J; submitted) There are two fundamentally different approaches to specifying and verifying properties of systems. The logical approach makes use of specifications given as formulae of temporal or modal logics and relies on efficient model checking algorithms; the behavioural approach exploits various equivalence or refinement checking methods, provided the specifications are given in the same formalism as implementations. In this paper we provide translations between the logical formalism of nu-calculus and the behavioural formalism of disjunctive modal transition systems. The translations preserve structural properties of the specification and allow us to perform logical operations on the behavioural specifications as well as behavioural compositions on logical formulae. The unification of both approaches provides additional methods for component-based stepwise design.

[4] (C) This paper studies a difference operator for stochastic systems whose specifications are represented by Abstract Probabilistic Automata (APAs). In the case refinement fails between two specifications, the target of this operator is to produce a specification APA that represents all witness PAs of this failure. Our contribution is an algorithm that allows to approximate the difference of two APAs with arbitrary precision. Our technique relies on new quantitative notions of distances between APAs used to assess convergence of the approximations, as well as on an in-depth inspection of the refinement relation for APAs. The procedure is effective and not more complex to implement than refinement checking.

[21] (C) We provide a framework for compositional and iterative design and verification of systems with quantitative information, such as rewards, time or energy. It is based on disjunctive modal transition systems where we allow actions to bear various types of quantitative information. Throughout the design process the actions can be further refined and the information made more precise. We show how to compute the results of standard operations on the systems, including the quotient (residual), which has not been previously considered for quantitative non-deterministic systems. Our quantitative framework has close connections to the modal nu-calculus and is compositional with respect to general notions of distances between systems and the standard operations.

[35] (J; submitted) We provide a framework for compositional and iterative design and verification of systems with quantitative information, such as rewards, time or energy. It is based on disjunctive modal transition systems where we allow actions to bear various types of quantitative information. Throughout the design process the actions can be further refined and the information made more precise. We show how to compute the results of standard operations on the systems, including the quotient (residual), which has not been previously considered for quantitative non-deterministic systems. Our quantitative framework has close connections to the modal nu-calculus and is compositional with respect to general notions of distances between systems and the standard operations.

[6] (J) This paper proposes a new theory of quantitative specifications. It generalizes the notions of stepwise refinement and compositional design operations from the Boolean to an arbitrary quantitative setting. Using a great number of examples, it is shown that this general approach permits to unify many interesting quantitative approaches to system design.

[7] (J) We present a distance-agnostic approach to quantitative verification. Taking as input an unspecified distance on system traces, or executions, we develop a game-based framework which allows us to define a spectrum of different interesting system distances corresponding to the given trace distance. Thus we extend the classic linear-time–branching-time spectrum to a quantitative setting, parametrized by trace distance. We also prove a general transfer principle which allows us to transfer counterexamples from the qualitative to the quantitative setting, showing that all system distances are mutually topologically inequivalent.

[25] (C) We introduce a new notion of structural refinement, a sound abstraction of logical implication, for the modal nu-calculus. Using new translations between the modal nu-calculus and disjunctive modal transition systems, we show that these two specification formalisms are structurally equivalent.Using our translations, we also transfer the structural operations of composition and quotient from disjunctive modal transition systems to the modal nu-calculus. This shows that the modal nu-calculus supports composition and decomposition of specifications.

### *6.5.2. Application papers:*

[32] (C; submitted) We suggest a method for measuring the degree to which features interact in feature-oriented software development. We argue that our method is practically feasible, easily extendable and useful from a developer's point of view.

[19] (C) Class diagrams are among the most popular modeling languages in industrial use. In a model-driven development process, class diagrams evolve, so it is important to be able to assess differences between revisions, as well as to propagate differences using suitable merge operations. Existing differencing and merging methods are mainly syntactic, concentrating on edit operations applied to model elements, or they are based on sampling: enumerating some examples of instances which characterize the difference between two diagrams. This paper presents the first known (to the best of our knowledge) automatic model merging and differencing operators supported by a formal semantic theory guaranteeing that they are semantically sound. All instances of the merge of a model and its difference with another model are automatically instances of the second model. The differences we synthesize are represented using class diagram notation (not edits, or instances), which allows creation of a simple yet flexible algebra for diffing and merging. It also allows presenting changes comprehensively, in a notation already known to users.

[20] (C) We propose a new similarity measure between texts which, contrary to the current state-of-the-art approaches, takes a global view of the texts to be compared. We have implemented a tool to compute our textual distance and conducted experiments on several corpuses of texts. The experiments show that our methods can reliably identify different global types of texts.

[23] (C) Reliable model transformations are essential for agile modeling. We propose to employ a configurable-semantics approach to develop automatic model transformations which are correct by design and can be integrated smoothly into existing tools and work flows.

[39] (C; submitted) Nowadays, large software systems are mostly built using existing services. These are not always designed to interact, i.e., their public interfaces often present some mismatches. Checking compatibility of service interfaces allows one to avoid erroneous executions when composing the services and ensures correct reuse and interaction. Service compatibility has been intensively studied, in particular for discovery purposes, but most of existing approaches return a Boolean result. In this paper, we present a quantitative approach for measuring the compatibility degree of service interfaces. Our method is generic and flooding-based, and fully automated by a prototype tool.

### *6.5.3. Surveys:*

[22] Modal transition systems provide a behavioral and compositional specification formalism for reactive systems. We survey two extensions of modal transition systems: parametric modal transition systems for specifications with parameters, and weighted modal transition systems for quantitative specifications.

[24] We survey extensions of modal transition systems to specification theories for probabilistic and timed systems.

## 6.6. Privacy and Security

**Participants:** Axel Legay, Fabrizio Biondi, Jean Quilbeuf, Thomas Given-Wilson.

### *6.6.1. Information-Theoretical Quantification of Security Properties*

This part of the work was not foreseen at the beginning of the action. It concerns security aspects, and more precisely quantifying privacy of data. This aspect is in fact central for SoS and all our algorithms developed for Tasks 4 and 5 should be adapted to solve a series of problems linked to privacy in interconnected object and dynamical environment. For now, we only studied the foundations.

Information theory provides a powerful quantitative approach to measuring security and privacy properties of systems. By measuring the *information leakage* of a system security properties can be quantified, validated, or falsified. When security concerns are non-binary, information theoretic measures can quantify exactly how much information is leaked. The knowledge of such informations is strategic in the developments of component-based systems.

The quantitative information-theoretical approach to security models the correlation between the secret information of the system and the output that the system produces. Such output can be observed by the attacker, and the attacker tries to infer the value of the secret by combining this information with its knowledge of the system.

Armed with the produced output and the source code of the system, the attacker tries to infer the value of the secret. The quantitative analysis we implement computes with arbitrary precision the number of bits of the secret that the attacker will expectedly infer. This expected number of bits is the information leakage of the system.

The quantitative approach generalizes the qualitative approach and thus provides superior analysis. In particular, a system respects non-interference if and only if its leakage is equal to zero. In practice very few systems respect non-interference, and for those who don't it is imperative to be able to distinguish between the ones leaking a very small amount of bits and the ones leaking a significant amount of bits, since only the latter are considered to pose a security vulnerability to the system.

Since black box security analyzes are immediately invalidated whenever an attacker gains information about the source code of the system, we assume that the attacker has a white box view of the system, meaning that it has access to the system's source code. This approach is also consistent with the fact that many security protocol implementations are in fact open source.

The scope of modern software projects is too large to be analyzed manually. For this reason we provide tools that can support the analyst and locate security vulnerabilities in large codebases and projects. We work with a variety of tools, including commercial software analysis tools being adapted with our techniques, and tools such as QUAIL developed here by our team.

We applied the leakage analysis provided by QUAIL to several case studies. Our case studies (voting protocol and smart grid coordination) have in common that a publicly disclosed information is computed from the secret of every participant in the model. In the voting example, the vote of a given voter is secret, but the number of votes for each candidates is public. Similarly, in the smart grid example, the consumption of one of the houses is secret, but the consumption of a whole quarter can be deduced. Qualitative analyses are either too restrictive or too permissive on these types of systems. For instance, non-interference will reject them as the public information depends on the secret. Declassification approaches will accept them, even if the number of voters or consumers is 2, in which case the secret can be deduced.

The development of better tools for quantitative security builds upon both theoretical developments in information theory, and development of the tools themselves. These often progress in parallel with each supporting the findings of the other, and increasing the demands and understanding upon each other.

*6.6.1.1. Papers:*

[3] (J; submitted) The quantification of information leakage provides a quantitative evaluation of the security of a system. We propose the usage of Markovian processes to model deterministic and probabilistic systems. By using a methodology generalizing the lattice of information approach we model refined attackers capable to observe the internal behavior of the system, and quantify the information leakage of such systems. We also use our method to obtain an algorithm for the computation of channel capacity from our Markovian models. Finally, we show how to use the method to analyze timed and non-timed attacks on the Onion Routing protocol.

[46] (C) Quantitative security analysis evaluates and compares how effectively a system protects its secret data. We introduce QUAIL, the first tool able to perform an arbitrary-precision quantitative analysis of the security of a system depending on private information. QUAIL builds a Markov Chain model of the system's behavior as observed by an attacker, and computes the correlation between

the system's observable output and the behavior depending on the private information, obtaining the expected amount of bits of the secret that the attacker will infer by observing the system. QUAIL is able to evaluate the safety of randomized protocols depending on secret data, allowing to verify a security protocol's effectiveness. We experiment with a few examples and show that QUAIL's security analysis is more accurate and revealing than results of other tools.

[40] (C; submitted) Quantitative security techniques have been proven effective to measure the security of systems against various types of attackers. However, such techniques are based on computing exponentially large channel matrices or Markov chains, making them impractical for large programs. We propose a different approach based on abstract trace analysis. By analyzing directly sets of execution traces of the program and computing security measures on the results, we are able to scale down the exponential cost of the problem. Also, we are able to appy statistical simulation techniques, allowing us to obtain significant results even without exploring the full space of traces. We have implemented the resulting algorithms in the QUAIL tool. We compare their effectiveness against the state of the art LeakWatch tool on two case studies: privacy of user consumption in smart grid systems and anonymity of voters in different voting schemes.

[12] (C) In an election, it is imperative that the vote of the single voters remain anonymous and undisclosed. Alas, modern anonymity approaches acknowledge that there is an unavoidable leak of anonymity just by publishing data related to the secret, like the election's result. Information theory is applied to quantify this leak and ascertain that it remains below an acceptable threshold. We apply modern quantitative anonymity analysis techniques via the state-of-the-art QUAIL tool to the voting scenario. We consider different voting typologies and establish which are more effective in protecting the voter's privacy. We further demonstrate the effectiveness of the protocols in protecting the privacy of the single voters, deriving an important desirable property of protocols depending on composite secrets.

[13] (C) In recent years, quantitative security techniques have been providing effective measures of the security of a system against an attacker. Such techniques usually assume that the system produces a finite amount of observations based on a finite amount of secret bits and terminates, and the attack is based on these observations. By modeling systems with Markov chains, we are able to measure the effectiveness of attacks on non-terminating systems. Such systems do not necessarily produce a finite amount of output and are not necessarily based on a finite amount of secret bits. We provide characterizations and algorithms to define meaningful measures of security for non-terminating systems, and to compute them when possible. We also study the bounded versions of the problems, and show examples of non-terminating programs and how their effectiveness in protecting their secret can be measured.

**FLUMINANCE Project-Team**

# 6. New Results

## 6.1. Highlights of the Year

### *6.1.1. Stochastic fluid flow dynamics under uncertainty*

We have proposed the basis of a formalism allowing to built large scale stochastic representation of fluid flows dynamics [17]. This formalism relies on a location uncertainty principle which separates the flow in terms of a resolved large scale component and a highly oscillating random component. The dynamics is built in a similar way as in the deterministic case through a stochastic representation of the Reynolds transport theorem. This principle paves a new way for the construction of subgrid models from the uncertainties we have on the flow. The associated subgrid tensor provides a clear interaction between small scale data and large scale resolved quantities. This characteristic opens new directions for the devising of methods for the nulmerical simulation of large scale components of the flow. It allows also deriving large-scale models that takes into account explicitly the inherent errors to a particular geophysical dynamics representation.

## 6.2. Fluid motion estimation

### *6.2.1. Stochastic uncertainty models for motion estimation*

**Participants:** Etienne Mémin, Manuel Saunier, Abed Malti.

In this study we have proposed a stochastic formulation of the brightness consistency used principally in motion estimation problems. In this formalization the image luminance is modeled as a continuous function transported by a flow known only up to some uncertainties. Stochastic calculus then enables to built conservation principles which take into account the motion uncertainties. These uncertainties defined either from isotropic or anisotropic models can be estimated jointly to the motion estimates. Such a formulation besides providing estimates of the velocity field and of its associated uncertainties allows us to naturally define a linear multiresolution scale-space framework. The corresponding estimator, implemented within a local least squares approach, has shown to improve significantly the results of the corresponding deterministic estimator (Lucas and Kanade estimator). This fast local motion estimator provides results that are of the same order of accuracy than state-of-the-art dense fluid flow motion estimator for particle images. The uncertainties estimated supply a useful piece of information in the context of data assimilation. This ability has been exploited to define multiscale incremental data assimilation filtering schemes. The development of an efficient GPU based version of this estimator recently started through the Inria ADT project FLUMILAB

### *6.2.2. 3D flows reconstruction from image data*

**Participants:** Ioana Barbu, Kai Berger, Cédric Herzet, 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. On the one hand, we investigate state-of-the-art methodologies such as ,“sparse representations" for the characterization of the observation and fluid motion models. On the other hand, we place the estimation problem into a probabilistic Bayesian framework and use state-of- the-art inference tools to effectively exploit the strong time-dependence on the fluid motion.

Last year, we focused on the design of new methodologies to jointly estimate the volume of particles and the velocity field from the received image data. Our approach was based on the minimization (with respect to both the position of the particles and the velocity field) of a cost function penalizing both the discrepancies with respect to a conservation equation and some prior estimates of particle positions.

This year, we revisited the problem of volume reconstruction through the prism of some modern optimization techniques. More specifically, we focussed our attention on the family of proximal and splitting methods and showed that the standard techniques commonly adopted in the TomoPIV literature can be seen as particular cases of such methodologies. Recasting standard methodologies in a more general framework allowed us to propose extensions of the latter: i) we showed that the parcimony characterizing the sought volume can be accounted for without increasing the complexity of the algorithms (e.g., by including simple thresholding operations); ii) we emphasized that the speed of convergence of the standard reconstruction algorithms can be improved by using Nesterov's acceleration schemes; iii) we also proposed a totally novel way of reconstructing the volume by using the so-called "alternating direction of multipliers method" (ADMM) . The journal publications relative to the contributions developped this year are currently in construction.

### *6.2.3. Sparse-representation algorithms*

**Participant:** Cédric Herzet.

The paradigm of sparse representations is a rather new concept which turns out to be central in many domains of signal processing. In particular, in the field of fluid motion estimation, sparse representation appears to be potentially useful at several levels: i) it provides a relevant model for the characterization of the velocity field in some scenarios; ii) it plays a crucial role in the recovery of volumes of particles in the 3D Tomo-PIV problem.

Unfortunately, the standard sparse representation problem is known to be NP hard. Therefore, heuristic procedures have to be devised to access to the solution of this problem. Among the popular methods available in the literature, one can mention orthogonal matching pursuit (OMP), orthogonal least squares (OLS) and the family of procedures based on the minimization of $\ell_p$ norms. In order to assess and improve the performance of these algorithms, theoretical works have been undertaken in order to understand under which conditions these procedures can succeed in recovering the "true" sparse vector.

Last, we contributed to this research axis by deriving conditions of success for the algorithms mentioned above when some partial information is available about the position of the nonzero coefficients in the sparse vector. This paradigm is of interest in the Tomographic-PIV volume reconstruction problem: one can indeed expect volumes of particles at two successive instants to be pretty similar; any estimate of the position of the particles at one given instant can therefore serve as a prior estimate about their position at the next instant. Another information of interest which can help the algorithms in their reconstruction process is the decay of the amplitude of the nonzero coeffcient in the sparse vector. In a TomoPIV context, this decay corresponds to the fact that not all the particles in fluid diffuse the same quantity of light (notably beacuse of illumination or radius variation). This year, we thus pursue our effort in the understanding of the success of some reconstruction algorithms when the sparse vectors obey some decay. In particular, we showed that the standard coherence-based guarantees for OMP/OLS can be relaxed by an amount which depends on the decay of the nonzero coeffcients.

Another axis of research we have dealt with is the extension of sparse methodologies to the context of nonlinear models. This type of situtation is indeed frequently encountered in fluid mechanics or geophysics where the initial/boundary conditions of a system are known to be sparse in some basis and the collected observations obey a nonlinear dynamical model (e.g., the Navier-Stokes equations). In our work, we showed that many sparse representation algorithms, designed in the linear paradigm, can be nicely extended to the nonlinear setup provided that the gradient of the functional can be evaluated efficiently. In order to do so, we suggested a methodology, well-known in the commmunity of optimal control, but surprinsingly quite uncommon in many fields of signal processing.

Our work have led to the publication of contributions in the IEEE International Conference on Speech, Acoustic and Signal Processing (ICASSP) [23] and international - Traveling Workshop on Interactions between Sparse models and Technology (iTwist) [22],[24]

## 6.3. Tracking, Data assimilation and model-data coupling

### 6.3.1. Stochastic filtering technique for the tracking of closed curves

**Participant:** Etienne Mémin.

We have studied a stochastic filtering technique for the tracking of closed curves along an image sequence. In that goal, we designed a continuous-time stochastic dynamics that allows us to infer inter-frame deformations. The curve is defined by an implicit level-set representation and the stochastic dynamics is expressed properly on the level-set function. It takes the form of a stochastic partial differential equation with a Brownian motion of low dimension. The evolution model we proposed combines local photometric information, deformations induced by the curve displacement and an uncertainty modeling of the dynamics. Specific choices of noise models and drift terms lead to an evolution law based on mean curvature as in classic level set methods, while other choices yield new evolution laws. The approach we propose is implemented through a particle filter, which includes color measurements characterizing the target and the background photometric probability densities respectively. The merit of this parameter free filter is demonstrated on various satellite image sequences depicting the evolution of complex geophysical flows. This work has been recently published in the Journal of Mathematical Imaging and Vision [13]. Let us note the method provides an empirical dynamical model learned recursively from a data flow. Its short time forecasting skills have been used in the context of weather-watch radar images within a fruitful collaboration with MeteoFrance.

### 6.3.2. Sequential smoothing for fluid motion

**Participants:** Anne Cuzol, Etienne Mémin.

In parallel to the construction of stochastic filtering techniques for fluid motions, we have proposed a new sequential smoothing method within a Monte-Carlo framework. This smoothing aims at reducing the temporal discontinuities induced by the sequential assimilation of discrete time data into continuous time dynamical models. The time step between observations can indeed be long in environmental applications for instance, and much longer than the time step used to discretize the model equations. While the filtering aims at estimating the state of the system at observations times in an optimal way, the objective of the smoothing is to improve the estimation of the hidden state between observation times. The method is based on a Monte-Carlo approximation of the filtering and smoothing distributions, and relies on a simulation technique of conditional diffusions. The proposed smoother can be applied to general non linear and multidimensional models. It has been applied to a turbulent flow in a high-dimensional context, in order to smooth the filtering results obtained from a particle filter with a proposal density built from an Ensemble Kalman procedure. This conditional simulation framework can also be used for filtering problem with low measurement noise. This has been explored through a collaboration with Jean-Louis Marchand (ENS Bretagne) in the context of vorticity tracking from image data.

### 6.3.3. Stochastic fluid flow dynamics under uncertainty

**Participants:** Etienne Mémin, Valentin Resseguier.

In this research axis we aim at devising Eulerian expressions for the description of fluid flow evolution laws under uncertainties. Such an uncertainty is modeled through the introduction of a random term that allows taking into account large-scale approximations or truncation effects performed within the dynamics analytical constitution steps. This includes for instance the modeling of unresolved scales interaction in large eddies simulation (LES) or in Reynolds average numerical simulation (RANS), but also uncertainties attached to non-uniform grid discretization. This model is mainly based on a stochastic version of the Reynolds transport theorem. Within this framework various simple expressions of the drift component can be exhibited for different models of the random field carrying the uncertainties we have on the flow. We aim at using such a formalization within image-based data assimilation framework and to derive appropriate stochastic versions of geophysical flow dynamical modeling. This formalization has been published in the journal Geophysical and Astrophysical Fluid Dynamics [17]. Numerical simulation on divergence free wavelets basis of 3D viscous Taylor-Green vortex and Crow instability have been performed within a collaboration with Souleymane Kadri-Harouna. First promising results have been obtained. Besides, we explore in the context of Valentin Resseguier's PhD the extension of such framework to oceanic models and to satellite image data assimilation. This PhD thesis takes place within a fruitful collaboration with Bertrand Chapron (CERSAT/IFREMER).

### 6.3.4. Free surface flows reconstruction and tracking

**Participants:** Dominique Heitz, Etienne Mémin.

We investigated the combined use of a Kinect depth sensor and of a stochastic data assimilation method to recover free-surface flows. More generally, we proposed a particle filter method to reconstruct the complete state of free-surface flows from a sequence of depth images only. The data assimilation scheme introduced accounts for model and observations errors. We evaluated the developed approach on two numerical test cases: a collapse of a water column as a toy-example and a flow in an suddenly expanding flume as a more realistic flow. The robustness of the method to simulated depth data quality and also to initial conditions was considered. We illustrated the interest of using two observations instead of one observation into the correction step. Then, the performance of the Kinect sensor to capture temporal sequences of depth observations was investigated. Finally, the efficiency of the algorithm was qualified for a wave in a real rectangular flat bottom tank. It was shown that for basic initial conditions, the particle filter rapidly and remarkably reconstructed velocity and height of the free surface flow based on noisy measures of the elevation alone. These results have been recently submitted to a special issue of Fluid Dynamics Research.

### 6.3.5. Variationnal ensemble methods for data assimilation

**Participants:** Dominique Heitz, Etienne Mémin, Cordelia Robinson, Yin Yang.

In this work, we aim at studying an ensemble based optimal control strategy for data assimilation. Such a formulation nicely combines the ingredients of ensemble Kalman filters and variational assimilation. In the same way as standard variational assimilation, it is formulated as the minimization of an objective function. However, similarly to ensemble filters, it introduces in its objective function an empirical ensemble-based background-error covariance and works in an off-line smoothing mode rather than sequentially like filtering approaches in a sequential filter. These techniques have the great advantage to avoid the constitution of tangent linear and adjoint models, which are necessary for standard incremental variational techniques. As the background error covariance matrix plays a key role in the variational process, our study particularly focuses on the generation of the analysis ensemble state with localization techniques. The proposed method was assessed with a Shallow Water model combined with synthetic data and original incomplete experimental depth sensor observations. Results submitted to Computers & Fluids showed that the modified ensemble technique was better in quality and reduced the computational cost.

### 6.3.6. Optimal control techniques for the coupling of large scale dynamical systems and image data

**Participants:** Dominique Heitz, Etienne Mémin, Cordelia Robinson.

This work aims at investigating the use of optimal control techniques for the coupling of Large Eddies Simulation (LES) techniques and 2D image data. The objective is to reconstruct a 3D flow from a set of simultaneous time resolved 2D image sequences visualizing the flow on a set of 2D plans enlightened with laser sheets. This approach will be experimented on shear layer flows and on wake flows generated on the wind tunnel of Irstea Rennes. Within this study we wish also 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 or in longer-term goal to learn empirical subgrid models directly from image data. An accurate modeling of this term is essential for Large Eddies Simulation as it models all the non resolved motion scales and their interactions with the large scales.

We have pursued the first investigations on a 4DVar assimilation technique, integrating PIV data and Direct Numerical Simulation (DNS), to reconstruct two-dimensional turbulent flows. The problem we are dealing with consists in recovering a flow obeying Navier-Stokes equations, given some noisy and possibly incomplete PIV measurements of the flow. By modifying the initial and inflow conditions of the system, the proposed method reconstructs the flow on the basis of a DNS model and noisy measurements. The technique has been evaluated in the wake of a circular cylinder. It denoises the measurements and increases the spatiotemporal

resolution of PIV time series. These results have been recently published in the Journal of Computational Physics [6]. Along the same line of studies the 3D case is ongoing. The goal consists here to reconstruct a 3D flow from a set of simultaneous time resolved 2D images of planar sections of the 3D volume. This work is mainly conducted within the PhD of Cordelia Robinson. The development of the variational assimilation code has been initiated within a collaboration with A. Gronskis, S. Laizé (lecturer, Imperial College, UK) and Eric Lamballais (institut P' Poitiers). A High Reynolds number simulation of the wake behind a cylinder has been recently performed within this collaboration. The 4DVar assimilation technique based on the numerical code Incompact3D is now implemented. We are currently trying to reconstruct a 3D turbulent flow from dual plane velocity observations. First assessments have been carried out with DNS based synthetic data. Further evaluation will be done with real measurements based on dual stereo PIV experiments.

### *6.3.7. Ensemble variational data assimilation of large scale fluid flow dynamics with uncertainty*

**Participants:** Etienne Mémin, Yin Yang.

In this work we explore the assimilation of a large scale representation of the flow dynamics with image data of finer resolution. The velocity field at large scales is described as a regular smooth component whereas the complement component is a highly oscillating random velocity field defined on the image grid but living at all the scales. Following this route we have started to assess the performance of a variational assimilation technique with direct image data observation. Preliminary encouraging results have been obtained for simulation under uncertainty of 1D and 2D shallow water models.

### *6.3.8. Reduced-order models for flows representation from image data*

**Participants:** Cédric Herzet, Etienne Mémin.

One of the possibilities to neglect the influence of some degrees of freedom over the main characteristics of a flow consists in representing it as a sum of $K$ orthonormal spatial basis functions weighted with temporal coefficients. To determine the basis function of this expansion, one of the usual approaches relies on the Karhunen-Loeve decomposition (refered to as proper orthogonal decomposition – POD – in the fluid mechanics domain). In practice, the spatial basis functions, also called modes, are the eigenvectors of an empirical auto-correlation matrix which is built from "snapshots" of the considered physical process.

In this axis of work we focus on the case where one does not have a direct access to snapshots of the considered physical process. Instead, the POD has to be built from the partial and noisy observations of the physical phenomenon of interest. Instances of such scenarios include situations where real instantaneous vector-field snapshots are estimated from a sequence of images. We have been working on several approaches dealing with such a new paradigm. A first approach consists in extending standard penalized motion-estimation algorithms to the case where the sought velocity field is constrained to span a low-dimensional subspace. In particular, we have considered scenarios where the standard optical flow constraint (OFC) is no longer satisfied and one has therefore to resort to a Discrete Finite Difference (DFD) model. The non-linearity of the latter leads to several practical issues that we have addressed this year.

## 6.4. Analysis and modeling of turbulent flows

### *6.4.1. Hot-wire anemometry at low velocities*

**Participant:** Dominique Heitz.

A new dynamical calibration technique has been developed for hot-wire probes. The technique permits, in a short time range, the combined calibration of velocity, temperature and direction calibration of single and multiple hot-wire probes. The calibration and measurements uncertainties were modeled, simulated and controlled, in order to reduce their estimated values. Based on a market study the french patent application has been extended this year to a Patent Cooperation Treaty (PCT) application.

### *6.4.2. Numerical and experimental image and flow database*

**Participant:** 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 carried out orthogonal dual plane time resolved stereoscopic PIV measurements in turbulent flows. The diagnostic employed two orthogonal and synchronized stereoscopic PIV measurements to provide the three velocity components in planes perpendicular and parallel to the streamwise flow direction. These temporally resolved planar slices observations will be used in 4DVar assimilation technique, integrating Direct Numerical Simulation (DNS) and Large Eddies Simulation (LES), to reconstruct three-dimensional turbulent flows. This reconstruction will be conducted within the PhD of Cordelia Robinson. The third challenge was to carried out a time resolved tomoPIV experiments in a turbulent wake flow. These temporally resolved volumic observations will be used to assess the algorithms developped in the PhD of Ioana Barbu and in the postdoc of Kai Berger. Then this data will be used in 4DVar assimilation technique to reconstruct three-dimensional turbulent flows. This reconstruction will be conducted within the PhD of Cordelia Robinson.

## 6.5. Visual servoing approach for fluid flow control

### *6.5.1. Minimization of the kinetic energy density in the 2D plane Poiseuille flow*

**Participants:** Christophe Collewet, Xuan Quy Dao.

This works concerns the PhD thesis of Xuan-Quy Dao. This year we have focused on a way to ensure a strict decreasing of the kinetic energy density. In that purpose, we have first proposed an approach to increase the controlled degrees of freedom. Indeed, the classical way to model this flow leads to only two degrees of freedom. With so few degrees of freedom it is obviously impossible to reach high desired performances as the strict minimization of the kinetic energy density. This way to proceed leads to a better minimization of the kinetic energy density. We have also proposed on approach based on a local decoupling of the controlled degree of freedom of the system so that an exponential decoupled decrease of each components of the state vector is locally obtained.

### *6.5.2. Control of systems described by partial differential equations*

**Participants:** Tudor-Bogdan Airimitoaie, Christophe Collewet.

This work concerns principally the post-doctoral research of Tudor-Bogdan Airimiţoaie. It aims at controlling continuously evolving systems described by partial differential equations (PDEs). This is relevant in the context of the Fluminance team because fluid flows are infinite dimensional systems and can be rigorously described only through PDEs. In spite of this, practical approaches of flow control are based on low order numerical implementation relying on space and time discretization of the continuous system. This implies to setup strategies for model reduction that must be then in return properly understood with respect to the convergence of the control law. For finite dimensional implementations, one of the research directions pursued concerns the study on the benefit of increasing the controlled degrees of freedom (see the work of Xuan-Quy Dao). Another research direction, started recently, consists in improving control by using real-time estimation of a finite number of parameters related to the original infinite dimensional system. Indeed, this opens the possibility of improving performances by using more advanced robust linear parametric varying (LPV) control techniques existing in the literature. Two conference papers on these works have been submitted at the "7th AIAA Flow Control Conference".

**GENSCALE Project-Team**

# 6. New Results

## 6.1. Highlights of the Year

**discoSnp published in NAR**. The publication presents a wide range of discoSnp applications that highlight the advantages and the drawbacks of predicting SNPs when no reference genomes are available. The publication witnesses the enthusiasm of users regarding both reference-free methods and the quality of the method. [20]

## 6.2. NGS methodology

**Participants:** Erwan Drezen, Anaïs Gouin, Dominique Lavenier, Claire Lemaitre, Antoine Limasset, Pierre Peterlongo, Guillaume Rizk.

**Comparison of large sets of metagenomics data**
We significantly extend the previous method (implemented in the Comparead tool) for computing similarity between sets of raw non assembled (and usually non-assemblable with current state of the art assemblers) reads. This enhancement of the method enables computations to be factorized when N read sets have to be compared all together. Moreover, the great advantage of this improvment is to save huge disk space and to enable efficient logical operations between metagenomic subset of reads. The Commet tool implements this optimized version.[25]

**De novo SNP discovery**
We developed a very efficient new way for detecting isolated SNPs given one, two or more raw read set(s) without using any reference genome. The implementation, called discoSnp, was applied to various datasets and applications. In particular, compared to finding isolated SNPs using a state-of-the-art assembly and mapping approach, our method requires significantly less computational resources, shows similar precision/recall values, and highly ranked predictions are less likely to be false positives. An experimental validation was conducted on an arthropod species (the tick *Ixodes ricinus*) on which de novo sequencing was performed. Among the predicted SNPs that were tested, 96% were successfully genotyped and truly exhibited polymorphism. [20]

**De novo discovery of inversion breakpoints**
A formal model has been proposed, together with an algorithm, for detecting inversion breakpoints without a reference genome, directly from raw NGS data. This model is characterized by a fixed size topological pattern in the de Bruijn Graph. We describe precisely the possible sources of false positives and false negatives and we additionally propose a sequence-based filter giving a good trade-off between precision and recall of the method. We implemented these ideas in a software called TakeABreak. Applied on simulated inversions in genomes of various complexity (from E. coli to a human chromosome dataset), the method provided promising results with a low memory footprint and a small computational time. [24]

**Integrated detection and assembly of long insertion variants**
We investigated a new method for the integrated detection and assembly of insertion variants from re-sequencing data. Contrary to other tools, it is designed to call insertions of any size, whether they are novel or duplicated, homozygous or heterozygous in the donor genome. We uses an efficient k-mer based method to detect insertion sites in a reference genome, and subsequently assemble them from the complete set of donor reads. The method is implemented in the tool MindTheGap and showed high recall and precision on simulated datasets of various genome complexities. When applied to real *C. elegans* and human datasets, MindTheGap detected and correctly assembled insertions longer than 1 kb, using at most 14 GB of memory. [19], [40]

**Enhancement of de-Bruijn Graph data structure**

The data structure holding the de-Bruijn Graph at the core of the GATB library has been improved through several new developments. First, its construction time has been greatly decreased thanks to the use of minimizers for kmer-counting, and efficient parallelization of various construction steps. Secondly, exploration of the graph has also been made faster through the possibility of parallel enumeration of nodes of interest, and through the use of a cache-coherent (blocked) bloom filter. Lastly, the structure itself has been extended to optionally allow for more information to be held, at a reasonable memory cost. A minimal perfect hash function allows to store additional data for each node, for example the coverage of each kmer. [11], [35], [36]

**Chloroplast assembly**
When sequencing plants, reads that correspond to the chloroplast genome are often over-represented. Filtering these reads based on k-mer counts allows specific assembly of the chloroplast to be directly performed. The small number of contigs can then be processed using advanced optimization tools to generate scaffolds. The approach has been partially tested on seqencing data from *Lactococcus lactis* to assemble plasmids of this bacteria. [12]

## 6.3. NGS applications

**Participants:** Susete Alves Carvalho, Rumen Andonov, Anaïs Gouin, Fabrice Legeai, Dominique Lavenier, Claire Lemaitre, Pierre Peterlongo, Ivaylo Petrov, Guillaume Rizk.

**Identification of genomic regions of biological interest**
The extraction and selection of 400 microsatellites among the large and fragmented *Acyrthosiphon pisum* genome led to the identification of a single 9cM region controlling the loss of sex in the pea aphid. The genotyping of these markers on geographically distant populations under divergent selection for reproductive strategies revealed a strong signature of selection in this genomic region, suggesting gene flow between populations with distinct reproductive modes. [15]

**Transcriptome assembly**
For this study, we incorporated various sources of RNA sequences from 454, Illumina and Sanger sequencing technologies obtained from more than 35 *S. frugiperda* developmental time-points and tissue samples and developed a custom pipeline to achieve their assembly. As a result, we provided a first valid transcriptome for *Spodoptera frugiperda*, a major agricultural pest. [16]

**Catalogue of long non coding RNAs**
We established a new bioinformatics pipeline for the detection of lncRNAs from RNA-Seq data, produced the first catalogue of aphids lncRNAs, and asserted for each lncRNA a classification of putative cis-interactions based on its genomic distance to neighboring mRNAs. These results allow the constitution of a broad gene regulation network of the aphid phenotypic plasticity at the embryo level. This workflow is available in Galaxy on the BioInformatics Platform for Arthopods of Agroecosystems (www.inra.fr/bipaa) and can be applied to any organism for which an annotated genome sequence and RNA-Seq data are provided.[23]

**Identification and correction of genome mis-assemblies due to heterozygosity**
Assembly tools are more and more efficient to reconstruct a genome from next-generation sequencing data but some problems remain. One of them corresponds to mis-assemblies due to heterozygosity (2 alleles instead of a consensus). Thus, we propose a strategy to detect and correct false duplications in assemblies based on several metrics: sequence similarity, matche length and average read coverage. Our method allows to decrease redundancy in the genome assembly, to improve the scaffolding and then to increase the N50 statistic by removal of one of the two alleles or joining of scaffolds by their extremities. This method was applied on the *Spodoptera frugiperda* genome.[39]

**Questioning the classical re-sequencing analyses approach**

Classical re-sequencing analyses are based on a first step of read mapping, then only mapped reads are taken into account in following analyses such as variant calling. We investigated the sources of unmapped reads in aphid re-sequencing data of 33 individuals, and we demonstrated that these reads contain valuable information that should not be discarded as usually done in such analyses. For instance, the analysis of the contigs obtained from assembling the unmapped reads led to recover some divergent genomic regions previously excluded from analysis and to discover putative novel sequences of *A. pisum* and its symbionts. We proposed strategies, based on assembly and re-mapping, to aid the capture and interpretation of this information.[14]

**Application of discoSnp on pea data**
The pea is a non-model organism with a large (4.5 GB) and complex genome which has not been sequenced yet. We compared, on the same set of low depth pea sequences, the SNPs generated by discoSnp with those obtained with a previous SNP discovery pipeline, and those generated using classical mapping approach combining Bowtie2 and GATK tools. [31]

## 6.4. HPC and parallelism

**Participants:** Rumen Andonov, Charles Deltel, Dominique Lavenier, François Moreews, Ivaylo Petrov.

**Workflows**
New tools are needed to enable the quick design and the intensive parallel execution of bioinformatics processes. Therefore, we propose a new Dataflow oriented workflow management system dedicated to intensive bioinformatics tasks. We worked on the interoperability of bioinformatics workflows using a model driven approach. Our results enable new import / export capabilities between multiple workflow management environments and incite to create a unique shared workflow model.[28]

**Graph processing : the All-Pairs Shortest Paths problem**
This research work anticipates the need of processing huge graphs that are results of intensive genomic sequence comparison (bank to bank processing). We proposed a new algorithm for solving the all-pairs shortest-path problem for planar graphs and graphs with small separators that exploits the massive on-chip parallelism available in today's Graphics Processing Units (GPUs). Our algorithm, based on the Floyd-War shall algorithm, has near optimal complexity in terms of the total number of operations, while its matrix-based structure is regular enough to allow for efficient parallel implementation on the GPUs. By applying a divide-and-conquer approach, we are able to make use of multi-node GPU clusters, resulting in more than an order of magnitude speedup over the fastest known Dijkstra-based GPU implementation and a two-fold speedup over a parallel Dijkstra-based CPU implementation.[27]

**Benchmark of Alignment Search Tools**
Comparing sequences is a daily task in bioinformatics and many software try to fulfill this need by proposing fast execution times and accurate results. Introducing a new software in this field requires to compare it to recognized tools with the help of well defined metrics. A set of quality metrics is proposed that enables a systematic approach for comparing alignment tools. These metrics have been implemented in a dedicated software, allowing to produce textual and graphical benchmark artifacts. [21]

## 6.5. Protein Structure

**Participants:** Rumen Andonov, Douglas Goncalves, Dominique Lavenier, Mathilde Le Boudic-Jamin, Antonio Mucherino.

**The molecular distance geometry problem**
The distance geometry is the problem of finding an embedding of a simple weighted undirected graph $G = (V, E, d)$ in a given dimension $K > 0$. Its most interesting application arises in biology, where the conformation of molecules such as proteins can be identified by embedding a graph (representing the molecular structure and some distance information) in dimension 3. Since some years, we are working on the discretization of the distance geometry. This year, the research developed in 4 main directions, that will be briefly detailed in the following paragraphs.

The majority of the work was performed on the so-called *discretization orders*, which are particular orders for the atoms of a molecule that allow for satisfying the discretization assumptions, i.e. they allow to discretize the search domain of the problem. Finding discretization orders is therefore an important pre-processing step for the solution of distance geometry problems. In fact, not only the identification of an atomic order allowing for the discretization is important, but also the identification of orders that are able to optimize some objectives that make the solution to the problem easier to perform. In this context, with both international and local partners, we worked on discretization orders that can be identified automatically in polynomial time [13], we worked on suitable orders for the protein side chains [10], and we studied some objectives to be optimized in discretization orders [38].

The algorithm that we mostly employ for the solution of distance geometry problems that can be discretized is the Branch & Prune (BP) algorithm. It recursively constructs the discretized search domain (a tree) and verifies the feasibility of the computed atomic positions. When all available distances are exact, all candidate positions for a given atom can be enumerated. This is however not possible in presence of interval distances, because a continuous subset of positions can actually be computed for the corresponding atoms. The focus of the work in [22] is on a new scheme for an adaptive generation of a discrete subset of candidate positions from this continuous subset. The generated candidate positions do not only satisfy the distances employed in the discretization process, but also additional distances that might be available (the so-called pruning distances).

Since the BP algorithm can loose in performance when dealing with large molecules containing several interval distances, we worked this year on a variation of the algorithm named BetaMDGP [29]. This is a work in collaboration with Korean researchers. The BetaMDGP algorithm is based on the concept of beta-complex, which is a geometric construct extracted from the quasi-triangulation derived from the Voronoi diagram of atoms.

From the theoretical side, we worked on two main directions. First, we proved that, in discretizable distance geometry problems where all available distances are exact, the total number of solutions is always a power of two. This is related to the fact that the discrete search space contains several symmetries [18]. Secondly, we tried to summarize in [37] the current issues for efficiently solving real-life instances of the distance geometry.

Finally, the work we performed during the last years, including another important results from other colleagues currently working on this topic, was summarized in an extensive survey on the discretization of the distance geometry [17].

**Distance measure between Protein structure**
We propose here a new distance measure for comparing two protein structures based on their contact map representations (CMO). This novel measure (max-CMO metric), satisfies all properties of a metric on the space of protein representations. Having a metric in that space allows to avoid pairwise comparisons on the entire database and thus to significantly accelerate exploring the protein space compared to non metric spaces. We show on a gold-standard classification benchmark sets that our exact k-nearest neighbor scheme classifies up to 95% and 99% of queries correctly. Our k-NN classification thus provides a promising approach for the automatic classification of protein structures based on contact map overlap. [26], [30]

**Local similarity of protein structure**
Finding similarities between protein structures is a main goal in molecular biology. Most of the existing tools preserve order and only find single alignments even when multiple similar regions exist. We propose a new seed-based approach that discovers multiple pairs of similar regions. Its computational complexity is polynomial and it comes with a quality guarantee that the returned alignments have both Root Mean Squared Deviations (coordinate-based as well as internal-distances based) lower than a given threshold, if such exists. We do not require the alignments to be order preserving, which makes our algorithm suitable for detecting similar domains when comparing multi-domain proteins. And because the search space for non-sequential alignments is much larger than for sequential ones, the computational burden is addressed by using both a coarse-grain level parallelism and a fine-grain level parallelism. [33]

**HYBRID Project-Team**

# 6. New Results

## 6.1. Highlights of the Year

- Paper [22] from Merwan Achibet, Maud Marchal, Ferran Argelaguet and Anatole Lécuyer received the "Best Paper Award" at IEEE Symposium on 3D User Interfaces 2014 (IEEE 3DUI'14).
- Paper [26] from Jean-Baptiste Barreau, Valérie Gouranton received the "Third Best Poster Award" at International Conference on Cultural Heritage 2014.

BEST PAPERS AWARDS :
[22] **IEEE Symposium on 3D User Interfaces**. M. ACHIBET, M. MARCHAL, F. ARGELAGUET SANZ, A. LÉCUYER.

## 6.2. 3D User Interfaces

### *6.2.1. 3D manipulation of virtual objects*

**Evaluation of Direct Manipulation using Finger Tracking for Complex Tasks in an Immersive Cube** Maud Marchal, Collaboration with REVES

We have proposed a solution for interaction using finger tracking in a cubic immersive virtual reality system (or immersive cube) [13]. Rather than using a traditional wand device, users can manipulate objects with fingers of both hands in a close-to-natural manner for moderately complex, general purpose tasks. Our solution couples finger tracking with a real-time physics engine, combined with a heuristic approach for hand manipulation, which is robust to tracker noise and simulation instabilities. A first study has been performed to evaluate our interface, with tasks involving complex manipulations, such as balancing objects while walking in the cube. The users finger-tracked manipulation was compared to manipulation with a 6 degree-of-freedom wand (or flystick), as well as with carrying out the same task in the real world. Users were also asked to perform a free task, allowing us to observe their perceived level of presence in the scene. Our results show that our approach provides a feasible interface for immersive cube environments and is perceived by users as being closer to the real experience compared to the wand. However, the wand outperforms direct manipulation in terms of speed and precision. We conclude with a discussion of the results and implications for further research

**A New Direct Manipulation Technique for Immersive 3D Virtual Environments** Thi Thuong Huyen Nguyen, Thierry Duval, Collaboration with MIMETIC

We have introduced a new 7-Handle manipulation technique [38] for 3D objects in immersive virtual environments and its evaluation. The 7-Handle technique includes a set of seven points which are flexibly attached to an object. There are three different control modes for these points including configuration, manipulation and locking/ unlocking modes. We have conducted an experiment to compare the efficiency of this technique with the traditional 6-DOF direct manipulation technique in terms of time, discomfort metrics and subjective estimation for precise manipulations in an immersive virtual environment in two consecutive phases: an approach phase and a refinement phase. The statistical results showed that the completion time in the approach phase of the 7-Handle technique was significantly longer than the completion time of the 6-DOF technique. Nevertheless, we found a significant interaction effect between the two factors (the manipulation technique and the object size) on the completion time of the refinement phase. In addition, even though we did not find any significant differences between the two techniques in terms of intuitiveness, ease of use and global preference in the result of subjective data, we obtained a significantly better satisfaction feedback from the subjects for the efficiency and fatigue criteria.

**A survey of plasticity in 3D user interfaces** Jérémy Lacoche, Thierry Duval, Bruno Arnaldi, Collaboration with b<>com

Plasticity of 3D user interfaces [33] refers to their capabilities to automatically fit to a set of hardware and environmental constraints. This area of research has already been deeply explored in the domain of traditional 2D user interfaces. Besides, during the last decade, interest in 3D user interfaces has grown. Designers find with 3D user interfaces new ways to promote and interact with data, such as e-commerce websites, scientific data visualization, etc. Because of the wide variety of Virtual Reality (VR) and Augmented Reality (AR) applications in terms of hardware, data and target users, there is a real interest in solutions for automatic adaption in order to improve the user experience in any context while reducing the development costs. An adaptation is performed in reaction to different criteria defining a system such as the targeted hardware platform, the user's context and the structure and the semantic of the manipulated data. This adaptation can then impact the system in different ways, especially content presentation, interaction techniques modifications and eventually the current distribution of the system across a set of available devices. In [33] we present the state of the art about plastic 3D user interfaces. Moreover, we present well known methods in the field of 2D user interfaces that could become relevant for 3D user interfaces.

### *6.2.2. Navigating in virtual environments*

**Adaptive Navigation in Virtual Environments** Ferran Argelaguet

Navigation speed for most navigation interfaces is still determined by rate-control devices (e.g. joystick). The interface designer is in charge of adjusting the range of optimal speeds according to the scale of the environment and the desired user experience. However, this approach is not valid for complex environments (e.g. multi-scale environments). Optimal speeds might vary for each section of the environment, leading to non-desired side effects such as collisions or simulator sickness. Thereby, we proposed a speed adaptation algorithm [24] based on the spatial relationship between the user and the environment and the user's perception of motion. The computed information is used to adjust the navigation speed in order to provide an optimal navigation speed and avoid collisions. Two main benefits of our approach is firstly, the ability to adapt the navigation speed in multi-scale environments and secondly, the capacity to provide a smooth navigation experience by decreasing the jerkiness of described trajectories. The evaluation showed that our approach provides comparable performance as existing navigation techniques but it significantly decreases the jerkiness of described trajectories

### *6.2.3. Novel pseudo-haptic based interfaces*

**Toward "Pseudo-Haptic Avatars": Modifying the Visual Animation ofSelf-Avatar Can Simulate the Perception of Weight Lifting**, Ferran Argelaguet, Anatole Lécuyer, Collaboration with MIMETIC

We have studied how the visual animation of a self-avatar can be artificially modified in real-time in order to generate different haptic perceptions [18]. In our experimental setup, participants could watch their self-avatar in a virtual environment in mirror mode while performing a weight lifting task. Users could map their gestures on the self-animated avatar in real-time using a Kinect. We introduce three kinds of modification of the visual animation of the self-avatar according to the effort delivered by the virtual avatar: 1) changes on the spatial mapping between the user's gestures and the avatar, 2) different motion profiles of the animation ,and 3) changes in the posture of the avatar (upper-body inclination). The experimental task consisted of a weight lifting task in which participants had to order four virtual dumbbells according to their virtual weight. The user had to lift each virtual dumbbells by means of a tangible stick, the animation of the avatar was modulated according to the virtual weight of the dumbbell. The results showed that the altering the spatial mapping delivered the best performance. Nevertheless, participants globally appreciated all the different visual effects. Our results pave the way to the exploitation of such novel techniques in various VR applications such as sport training, exercise games, or industrial training scenarios in single or collaborative mode.

**The Virtual Mitten: A Novel Interaction Paradigm for Visuo-Haptic Manipulation of Objects Using Grip Force** Merwan Achibet, Maud Marchal, Ferran Argelaguet, Anatole Lécuyer

We have proposed a novel visuo-haptic interaction paradigm called the "Virtual Mitten" [22] for simulating the 3D manipulation of objects. Our approach introduces an elastic handheld device that provides a passive haptic feedback through the fingers and a mitten interaction metaphor that enables to grasp and manipulate

objects. The grasping performed by the mitten is directly correlated with the grip force applied on the elastic device and a supplementary pseudo-haptic feedback modulates the visual feedback of the interaction in order to simulate different haptic perceptions. The Virtual Mitten allows natural interaction and grants users with an extended freedom of movement compared with rigid devices with limited workspaces. Our approach has been evaluated within two experiments focusing both on subjective appreciation and perception. Our results show that participants were able to well perceive different levels of effort during basic manipulation tasks thanks to our pseudo-haptic approach. They could also rapidly appreciate how to achieve different actions with the Virtual Mitten such as opening a drawer or pulling a lever. Taken together, our results suggest that our novel interaction paradigm could be used in a wide range of applications involving one or two-hand haptic manipulation such as virtual prototyping, virtual training or video game.

*Figure 2. The Virtual Mitten : Visuo-haptic manipulation as enabled by our novel approach called the "Virtual Mitten". Each hand holds an elastic device to control a virtual mitten (in grey) enabling the grasp of virtual objects.*

**Collaborative Pseudo-Haptics: Two-User Stiffness Discrimination Based on Visual Feedback** Ferran Argelaguet, Takuya Sato, Thierry Duval, Anatole Lécuyer, Collaboration with Tohoku University Research Institute of Electrical Communication

We have explored how the concept of pseudo-haptic feedback can be introduced in a collaborative scenario [25]. A remote collaborative scenario in which two users interact with a deformable object is presented. Each user, through touch-based input, is able to interact with a deformable virtual object displayed in a standard display screen. The visual deformation of the virtual object is driven by a pseudo-haptic approach taking into account both the user in-put and the simulated physical properties. Particularly, we investigated stiffness perception. In order to validate our approach, we tested our system in a single and two-user configuration. The results showed that users were able to discriminate the stiffness of the virtual object in both conditions with a comparable performance. Thus, pseudo-haptic feedback seems a promising tool for providing multiple users with physical information related to other users' interactions.

### *6.2.4. Sound and virtual reality*

**Sonic interaction with a virtual orchestra of factory machinery** Florian Nouviale, Valérie Gouranton, collaboration with Ronan Gaugne (IMMERSIA) and LIMSI

We have conceived an immersive application where users receive sound and visual feedbacks on their interactions with a virtual environment. In this application, the users play the part of conductors of an orchestra of factory machines since each of their actions on interaction devices triggers a pair of visual and audio responses. Audio stimuli were spatialized around the listener. The application was exhibited during the 2013 Science and Music day and designed to be used in a large immersive system with head tracking, shutter glasses and a 10.2 loudspeaker configuration [43].

**Audio-Visual Attractors for Capturing Attention to the Screens When Walking in CAVE Systems** Ferran Argelaguet, Valérie Gouranton, Anatole Lécuyer, collaboration with Aalborg University

In four-sided CAVE-like VR systems, the absence of the rear wall has been shown to decrease the level of immersion and can introduce breaks in presence. We have therefore investigated to which extent user's attention can be driven by visual and auditory stimuli in a four-sided CAVE-like system [32]. An experiment was conducted in order to analyze how user attention is diverted while physically walking in a virtual environment, when audio and/or visual attractors are present. The four sided CAVE used in the experiment allowed to walk up to 9m in straight line. An additional key feature in the experiment is the fact that auditory feedback was delivered through binaural audio rendering techniques via non-personalized head related transfer functions (HRTFs). The audio rendering was dependent on the user's head position and orientation, enabling localized sound rendering. The experiment analyzed how different "attractors" (audio and/or visual, static or dynamic) modify the user's attention. The results of the conducted experiment show that audio-visual attractors are the most efficient attractors in order to keep the user's attention toward the inside of the CAVE. The knowledge gathered in the experiment can provide guidelines to the design of virtual attractors in order to keep the attention of the user and avoid the "missing wall".

### *6.2.5. Experiencing the past in virtual reality*

**Immersia, an open immersive infrastructure: doing archaeology in virtual reality** Valérie Gouranton, Bruno Arnaldi, collaboration with MIMETIC and Ronan Gaugne (IMMERSIA)

We have first studied the mutual enrichment between archaeology and virtual reality [16]. To do so, we are considering Immersia, our open high-end platform dedicated to research on immersive virtual reality and its usages. Immersia is a node of the european project Visionair that offers an infrastructure for high level visualisation facilities open to research communities across Europe. In Immersia, two projects are currently active on the theme of archaeology. One is relative to the study of the Cairn of Carn, with the Creaah, a pluridisciplinary research laboratory of archeology and archeosciences, and one on the reconstitution of the gallo-roman villa of Bais, with the French institute INRAP.

**Virtual reality tools for the West Digital Conservatory of Archaeological Heritage** Jean-Baptiste Barreau, Valérie Gouranton, collaboration with Ronan Gaugne (IMMERSIA) and INRAP

In continuation of the 3D data production work made by the WDCAH (West Digital Conservatory of Archaeological Heritage), the use of virtual reality tools allows archaeologists to carry out analysis and understanding research about their sites. We have then focused on the virtual reality services proposed to archaeologists in the WDCAH, through the example of two archaeological sites, the Temple de Mars in Corseul and the Cairn of Carn Island [27].

**Preservative Approach to Study Encased Archaeological Artefacts** Valérie Gouranton, Bruno Arnaldi, collaboration with Ronan Gaugne (IMMERSIA) and INRAP

We have proposed a workflow based on a combination of computed tomography, 3D images and 3D printing to analyse different archaeological material dating from the Iron Age, a weight axis, a helical piece, and a fibula [39]. This workflow enables a preservative analysis of the artefacts that are unreachable because encased either in stone, corrosion or ashes. Computed tomography images together with 3D printing provide a rich toolbox for archaeologist work allowing to access a tangible representation of hidden artefacts. These technologies are combined in an efficient, affordable and accurate workflow compatible with preventive archaeology constraints.

**Combination of 3D Scanning, Modeling and Analyzing Methods around the Castle of Coatfrec Reconstitution** Jean Baptiste Barreau, Valérie Gouranton, collaboration with Ronan Gaugne (IMMERSIA) and INRAP

The castle of Coatfrec is a medieval castle in Brittany constituting merely a few remaining ruins currently in the process of restoration. Beyond its great archeological interest, it has become, over the course of the last few years, the subject of experimentation in digital archeology. Methods of 3D scanning were used in order to gauge comparisons between the remaining structures and their absent hypothetical ones, resulting in the first

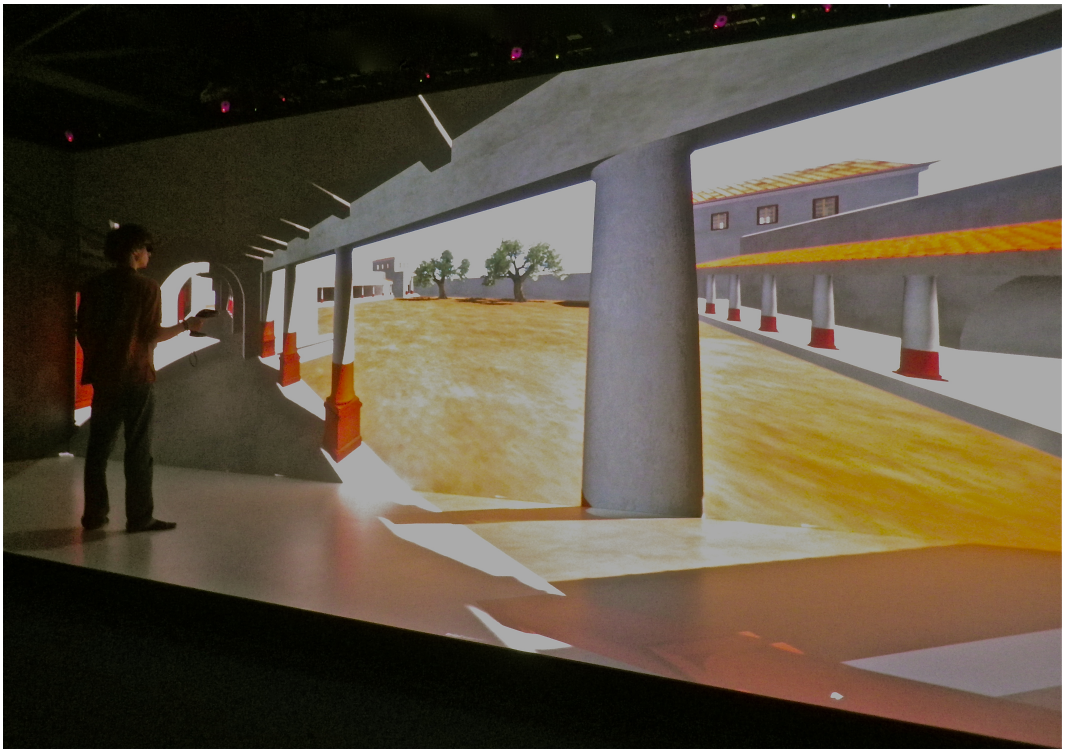

*Figure 3. Virtual visit of the "Temple of Mars" in the Immersia room.*

quantitative results of its kind. We have applied these methods and presented the subsequent results obtained using these new digital tools [26].

**Ceramics Fragments Digitization by Photogrammetry, Reconstructions and Applications** Jean Baptiste Barreau, Valérie Gouranton, collaboration with Ronan Gaugne (IMMERSIA) and INRAP

We have studied an application of photogrammetry on ceramic fragments from two excavation sites located north-west of France [28]. The restitution by photogrammetry of these different fragments allowed reconstructions of the potteries in their original state or at least to get to as close as possible. We used the 3D reconstructions to compute some metrics and to generate a presentation support by using a 3D printer. This work is based on affordable tools and illustrates how 3D technologies can be quite easily integrated in archaeology process with limited financial resources.

## 6.3. Physical simulation and multisensory feedback

### *6.3.1. Physically-based simulation and collision detection*

**Fast collision detection for fracturing rigid bodies** Loeiz Glondu, Maud Marchal

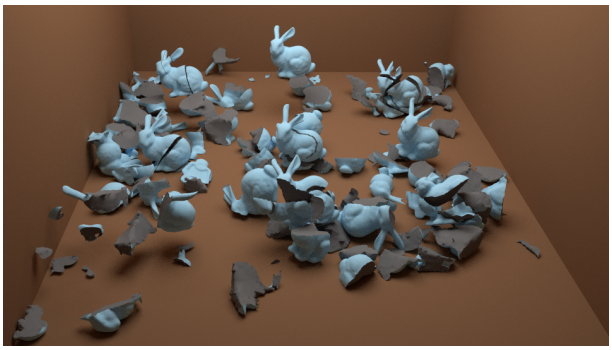

*Figure 4. Fast collision detection during real-time brittle fracture simulation.*

In complex scenes with many objects, collision detection plays a key role in the simulation performance. This is particularly true in fracture simulation for two main reasons. One is that fracture fragments tend to exhibit very intensive contact, and the other is that collision detection data structures for new fragments need to be computed on the fly. In [17], we present novel collision detection algorithms and data structures for real-time simulation of fracturing rigid bodies. We build on a combination of well-known efficient data structures, namely distance fields and sphere trees, making our algorithm easy to integrate on existing simulation engines. We propose novel methods to construct these data structures, such that they can be efficiently updated upon fracture events and integrated in a simple yet effective self-adapting contact selection algorithm. Altogether, we drastically reduce the cost of both collision detection and collision response. We have evaluated our global solution for collision detection on challenging scenarios, achieving high frame rates suited for hard real-time applications such as video games or haptics. Our solution opens promising perspectives for complex fracture simulations involving many dynamically created rigid objects.

This work was achieved in collaboration with Miguel Otaduy and Sara Schvartzman (URJC Madrid, Spain) and Georges Dumont (MIMETIC team).

**Collision detection: broad phase adaptation from multi-core to multi-GPU architecture** Bruno Arnaldi, Valérie Gouranton

We have presented several contributions on the collision detection optimization centered on hardware performance. We focus on the first step (Broad-phase) and propose three new ways of parallelization of the well-known Sweep and Prune algorithm. We first developed a multi-core model takes into account the number of available cores. Multi-core architecture enables us to distribute geometric computations with use of multi-threading. Critical writing section and threads idling have been minimized by introducing new data structures for each thread. Programming with directives, like OpenMP, appears to be a good compromise for code portability. We then proposed a new GPU-based algorithm also based on the "Sweep and Prune" that has been adapted to multi-GPU architectures. Our technique is based on a spatial subdivision method used to distribute computations among GPUs. Results show that significant speed-up can be obtained by passing from 1 to 4 GPUs in a large-scale environment [12].

**Real-time tracking of deformable target in ultrasound images** Maud Marchal

In several medical applications such as liver or kidney biopsies, an anatomical region needs to be continuously tracked during the intervention. When using ultrasound (US) image modality, tracking soft tissues remains challenging due to the deformations caused by physiological motions or medical instruments, combined with the generally weak quality of the images. In order to overcome the previous limitation, different techniques based on physical model have been proposed in the literature. In [41], we propose an approach for tracking deformable target within 2D US images based on a physical model driven by smooth displacement field obtained from dense information. This allows to take into account highly localized deformation in the US images.

This work was achieved in collaboration with Lucas Royer and Alexandre Krupa (Lagadic team), Anthony Le Bras (CHU Rennes) and Guillaume Dardenne (IRT B-Com).

### *6.3.2. Multimodal feedback*

**Stereoscopic Rendering of Virtual Environments with Wide Field-of-Views up to 360** Jérôme Ardouin, Anatole Lécuyer, Maud Marchal

We propose a novel approach [23] for stereoscopic rendering of virtual environments with a wide Field-of-View (FoV) up to 360. Handling such a wide FoV implies the use of non-planar projections and generates specific problems such as for rasterization and clipping of primitives. We propose a novel pre-clip stage specifically adapted to geometric approaches for which problems occur with polygons spanning across the projection discontinuities. Our approach integrates seamlessly with immersive virtual reality systems as it is compatible with stereoscopy, head-tracking, and multi-surface projections. The benchmarking of our approach with different hardware setups shows that it is well compliant with real-time constraints, and capable of displaying a wide range of FoVs. Thus, our geometric approach could be used in various VR applications in which the user needs to extend the FoV and apprehend more visual information.

This work was achieved in collaboration with Eric Marchand (Lagadic team).

**A survey on bimanual haptics** Anatole Lécuyer, Maud Marchal, Anthony Talvas.

When interacting with virtual objects through haptic devices, most of the time only one hand is involved. However, the increase of computational power, along with the decrease of device costs, allow more and more the use of dual haptic devices. The field which encompasses all studies of the haptic interaction with either remote or virtual environments using both hands of the same person is referred to as bimanual haptics. It differs from the common unimanual haptic field notably due to specificities of the human bimanual haptic system, e.g. the dominance of the hands, their differences in perception and their interactions at a cognitive level. These specificities call for adapted solutions in terms of hardware and software when applying the use of two hands to computer haptics. In [21], we review the state of the art on bimanual haptics, encompassing the human factors in bimanual haptic interaction, the currently available bimanual haptic devices, the software solutions for two-handed haptic interaction, and the existing interaction techniques.

**Haptic cinematography** Fabien Danieau, Anatole Lécuyer

Haptics, the technology which brings tactile or force-feedback to users, has a great potential for enhancing movies and could lead to new immersive experiences. In [14] we introduce *Haptic Cinematography* which presents haptics as a new component of the filmmaker's toolkit. We propose a taxonomy of haptic effects and we introduce novel effects coupled with classical cinematographic motions to enhance video viewing experience. More precisely we propose two models to render haptic effects based on camera motions: the first model makes the audience feel the motion of the camera and the second provides haptic metaphors related to the semantics of the camera effect. Results from a user study suggest that these new effects improve the quality of experience. Filmmakers may use this new way of creating haptic effects to propose new immersive audiovisual experiences.

This work was achieved in collaboration with Marc Christie (MIMETIC team), Julien Fleureau, Philippe Guillotel and Nicolas Mollet (Technicolor).

## 6.4. Collaborative Virtual Environments

### *6.4.1. Collaborative virtual environments for training*

**Collaborative Virtual Training with Physical and Communicative Autonomous Agents** Thomas Lopez, Valérie Gouranton, Florian Nouviale, Rozenn Bouville-Berthelot, Bruno Arnaldi

Virtual agents are a real asset in collaborative virtual environment for training (CVET) as they can replace missing team members. Collaboration between such agents and users, however, is generally limited. We presented a whole integrated model of CVET focusing on the abstraction of the real or virtual nature of the actor to define a homogenous collaboration model. First, we defined a new collaborative model of interaction. This model notably allows to abstract the real or virtual nature of a teammate. Moreover, we proposed a new role exchange approach so that actors can swap their roles during training. The model also permits the use of physically based objects and characters animation to increase the realism of the world. Second, we design a new communicative agent model, which aims at improving collaboration with other actors using dialog to coordinate their actions and to share their knowledge. Finally, we evaluated the proposed model to estimate the resulting benefits for the users and we show that this is integrated in existing CVET applications [20].

*Figure 5. Collaborative virtual environment for training (CVET) as experienced in our Immersia VR room*

**Exchange of avatars : Toward a better perception and understanding** Thomas Lopez, Rozenn Bouville-Berthelot, Florian Nouviale, Valérie Gouranton, Bruno Arnaldi

The exchange of avatars, i.e. the actual fact of changing once avatar with another one, is a promising trend in multi-actor virtual environments. It provides new opportunities for users, such as controlling a different avatar for a specific action, retrieving knowledge belonging to a particular avatar, solving conflicts and deadlocks situations or even helping another user. Virtual Environments for Training are especially affected by this trend as a specific role derived from a scenario is usually assigned to a unique avatar. Despite the increasing use of avatar exchange, users' perception and understanding of this mechanism have not been studied. We propose two complementary user-centered evaluations that aim at comparing several representations for the exchange of avatars; these are termed exchange metaphors. Our first experiment focuses on the perception of an exchange by a user who is not involved in the exchange, and the second experiment analyzes the perception of an exchange triggered by the user. Results show that the use of visual feedback globally aids better understanding of the exchange mechanism in both cases. Our first experiment suggests, however, that visual feedback is less efficient than a simple popup notification in terms of task duration. In addition, the second experiment shows that much simpler metaphors with no visual effect are generally preferred because of their efficiency [19].

*Figure 6. Example of metaphor used in "Exchange of Avatars": the blue expert exchanges avatar here with the red expert.*

**An interaction abstraction model for seamless avatar exchange in CVET** Rozenn Bouville-Berthelot, Thomas Lopez, Florian Nouviale, Valérie Gouranton, Bruno Arnaldi

Collaboration and interaction between users and virtual humans in virtual environments is a crucial challenge, notably for Collaborative Virtual Environments for Training (CVET). A training procedure, indeed, often involves several actors: trainees, teammates and many times a trainer. Yet, a major benefit of CVET is to propose to users to be trained even if the required number of person needed by the procedure is not available. Therefore, almost every CVET use autonomous virtual humans to replace the missing person. We have proposed to improve the effective collaboration between users and virtual humans involved in a complex

task within CVET. Using an entity called the "Shell", we are able to wrap the features common to both users and virtual humans. It gives us an abstraction level to pool the management of the main processes useful to control an avatar, interact with the environment and gather knowledge from a CVET. Besides, the Shell allows seamless exchange of avatars during a procedure. Thanks to the Shell, the exchange can be carried out at any time during a task while preserving all the data associated to a role in a procedure [29].

**#SEVEN: a Sensor Effector Based Scenarios Model for Driving Collaborative Virtual Environment** Guillaume Claude, Valérie Gouranton, Rozenn Bouville-Berthelot, Bruno Arnaldi

We introduced #SEVEN, a sensor effector model that enables the execution of complex scenarios for driving Virtual Reality applications. #SEVEN is based on an enhanced Petri net model which is able to describe and solve intricate event sequences. Our model also proposes several useful features for the design of collaborative scenarios for Collaborative Virtual Environments such as versatile roles and Activity Continuum. We also illustrate its usage it by describing a demonstrator that presents an implementation of our model [30].

**Collaborative virtual environments for ergonomics: embedding the design engineer role in the loop** Thierry Duval, collaboration with Charles Pontonnier and Georges Dumont (MIMETIC).

We have proposed to define the role and duties of a design engineer involved in a collaborative ergonomic design session supported by a 3D collaborative virtual environment. For example, such a session can be used to adapt the manual task an operator must achieve in the context of an industrial assembly line. We first presented the interest of such collaborative sessions. Then we presented a related work explaining the need of proper 3DCVE and metaphors to obtain efficient collaborative ergonomic design sessions. Then we proposed a use case highlighting the type of metaphor such engineers need to have to be efficient in such a framework [40].

### *6.4.2. Collaborative virtual environments and awareness*

**Improving Awareness for 3D Virtual Collaboration by Embedding the Features of Users' Physical Environments and by Augmenting Interaction Tools with Cognitive Feedback Cues** Thierry Duval, Thi Thuong Huyen Nguyen, Valérie Gouranton, collaboration with MimeTic

The feeling of presence is essential for efficient interaction within Virtual Environments (VEs). When a user is fully immersed within a VE through a large immersive display system, his/her feeling of presence can be altered because of disturbing interactions with his/her physical environment, such as collision with hardware parts of the system or loss of tracking. This alteration can be avoided by taking into account the physical features of the user as well as those of the system hardware and embedding them in the VE. Moreover, the 3D abstract representation of these physical features can also be useful for collaboration between distant users because they can make a user aware of the physical limitations of the others he/she is collaborating with. We used the Immersive Interactive Virtual Cabin (IIVC) model to obtain this virtual representation of the user's physical environment and we illustrated how this representation can be used in a collaborative navigation task in a VE. We also presented how we can add 3D representations of 2D interaction tools in order to cope with asymmetrical collaborative configurations, providing 3D cues for a user to understand the actions of the others even if he/she is not fully immersed in the shared VE [15].

**From 3D Bimanual Toward Distant Collaborative Interaction Techniques: An Awareness Issue** Morgan Le Chénechal, Thierry Duval, Valérie Gouranton, Bruno Arnaldi, collaboration with b<>com

CVE involve the use of complex interaction techniques based on specific collaborative metaphors. The design of these metaphors may be a difficult task because it has to deal with collaborative issues that came from sparse research areas (Human-Computer Interfaces, Human-Human Interactions, Networking, Physiology and Social Psychology). Metaphors for bimanual interactions have been developed for a while essentially because it is a widely spread area of interest for common tasks. Bimanual interactions involve the simultaneous use of both hands of the user in order to achieve a goal with better performances compared to uni-manual interactions thanks to a natural skill that is proprioception. This collaborative aspect could certainly be a helpful entry point in the design of efficient collaborative interaction techniques extended from improved bimanual metaphors. However, the proprioceptive sense cannot be considered in the same way, and additional features must be

proposed to be able to collaborate efficiently. Thus, awareness is a key to let CVE be usable and the availability of collaborative feedbacks is essential to extend bimanual interactions toward collaborative ones. In this paper, we based our study on existing work on bimanual and collaborative interaction techniques trying to draw similarities between them. We emphasize common points between both fields that could be useful to better design both metaphors and awareness in CVE [34].

**A survey of communication and awareness in collaborative virtual environments** Thi Thuong Huyen Nguyen, Thierry Duval

In the domain of Collaborative Virtual Environments (CVEs), many virtual worlds, frameworks and techniques are built based on a specific and direct purpose. There is not a general and still good and efficient enough solution for all the collaborative systems. Depending on the purpose of the collaborative work, the techniques of interaction and of manipulation change from one application to another. Despite this difference between interaction techniques, they always benefit greatly from awareness features that help in explicating implicit knowledge related to one's own and others' working activities as well as to virtual workspace. In addition, people in CVEs also use communication channels to negotiate shared understandings of task goals, of task decomposition and of task progress. Therefore, awareness and communication are usually considered as "instruments" to complete collaborative tasks in the environment. However, few research work have been devoted to improving the awareness and the communication channels in CVEs for a better collaboration between users. We have studied the importance of awareness and communication in collaborative virtual environments. We have investigated different kinds of awareness which need to be carefully designed. We have discussed different communication means and how to cope with this diversity, so we can benefit from the availability of different peripheral devices and can find an effective communication means to work together. Finally, we have made some propositions to overcome these actual limitations of CVEs [37].

### *6.4.3. Collaborative virtual environments and software engineering*

**When model driven engineering meets virtual reality: feedback from application to the Collaviz framework** Thierry Duval, collaboration with Arnaud Blouin and Jean-Marc Jézéquel (DIVERSE).

Despite the increasing use of 3D Collaborative Virtual Environments (3D CVE), their development is still a cumbersome task. The various concerns to consider (distributed system, 3D graphics, etc.) complexify the development as well as the evolution of CVEs. Software engineering recently proposed methods and tools to ease the development process of complex software systems. Among them, Model-Driven Engineering (MDE) considers models as first-class entities. A model is an abstraction of a specific aspect of the system under study for a specific purpose. MDE thus breaks down a complex system into as many models for different purposes, such as: generating code from models; building domain specific programming/modeling languages (DSL); generating tools such as graphical or textual editors. We have leveraged MDE for developing 3D CVEs. We showed how the Collaviz framework took benefits from a DSL we built. The benefits are multiple: 3D CVE designers can focus on the behavior of their virtual objects without bothering with distributed and graphics features; configuring the content of 3D CVEs and their deployment on various software and hardware platforms can be automated through code generation. We detailed the development process we propose and the experiments we conducted on Collaviz [31].

## 6.5. Brain-Computer Interfaces

### *6.5.1. Novel usages of BCI*

**Mind-Mirror: combining BCI and augmented reality to "see your brain in action in your own head"**, Anatole Lécuyer, Jonathan Mercier, Maud Marchal

Imagine you are facing a mirror, seeing at the same time both your real body and a virtual display of your brain in activity and perfectly superimposed to your real image "inside your real skull". We have introduced a novel augmented reality paradigm called "Mind-Mirror" which enables the experience of seeing "through your own head", visualizing your brain "in action and in situ" [36]. Our approach relies on the use of a semi-transparent mirror positioned in front of a computer screen. A virtual brain is displayed on screen

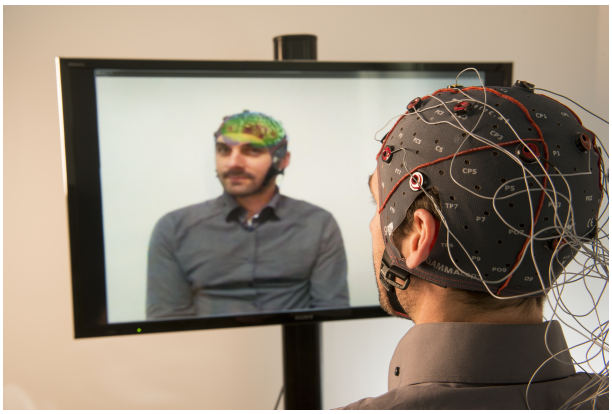

*Figure 7. The Mind Mirror system : Visualization of brain activity in real-time using a mirror-based augmented reality setup and an EEG headset.*

and automatically follows the head movements using an optical face-tracking system. The brain activity is extracted and processed in real-time with the help of an electroencephalography cap (EEG) worn by the user. A rear view is also proposed thanks to an additional webcam recording the rear of the user's head. The use of EEG classification techniques enables to test a Neurofeedback scenario in which the user can train and progressively learn how to control different mental states, such as "concentrated" versus "relaxed". The results of a user study comparing a standard visualization used in Neurofeedback to our approach showed that the Mind-Mirror could be successfully used and that the participants have particularly appreciated its innovation and originality. We believe that, in addition to applications in Neurofeedback and Brain-Computer Interfaces, the Mind-Mirror could also be used as a novel visualization tool for education, training or entertainment applications.

This work was achieved in collaboration with Fabien Lotte from POTIOC team (Inria-Bordeaux).

**Using SSVEP-based BCI with 3D stereoscopic display** Anatole Lécuyer

We have investigated the feasibility of dual-frequency Steady-State Visual Evoked Potential (SSVEP) stimulation using a 3-D display and stereoscopic glasses [44]. Dual-frequency stimulation allows for more targets to be created using a small number of frequencies, and stereoscopic vision offers a suitable medium for dual-frequency stimulation as the two views can be controlled independently. Participants were exposed to a repetitive visual stimulus flashing at different frequencies in the left and right views and the electroencephalography (EEG) trace was examined. Our results suggest that the two stimulation frequencies can still be evident in the SSVEP response. In addition, the participant ratings showed no significant differences in fatigue, annoyance, comfort or strangeness of the stimulation compared to conventional forms of stimulation. These results pave the way for further studies using stereoscopic dual-frequency stimulation and its potential for use in virtual reality and 3D videogames

This work was achieved in collaboration with Robert Leeb (EPFL, Switzerland).

**Passive BCI and music** Anatole Lécuyer

Passive brain–computer interfaces (passive BCI), also named implicit BCI, provide information from user mental activity to a computerized application without the need for the user to control his brain activity. We have proposed an overview of current research on passive BCIs in [45]. We have notably studied how they can be applied to the context of music creation, where they can provide novel information to adapt the music creation process, e.g., exploiting user mental concentration to adapt the music tempo.

### 6.5.2. BCI methodology

**Which factors drive successful BCI skill learning?** Anatole Lécuyer, Lorraine Perronnet

Brain-Computer Interfaces although very promising, suffer from a poor reliability. Rather than improving brain signal-processing alone, an interesting research direction is to guide users to learn BCI control mastery. Thus, we have inroduced a set of motivational and cognitive factors which could influence the learning process, and which should be considered to improve the global performance of BCI users [47]. We base our approach on Keller's integrative theory of motivation, volition, and performance, which combines motivational (affective) and cognitive factors, to explain what makes human users learn and perform efficiently, irrespectively of the task. These factors can guide the creation of learning environments, such as BCI training protocols.

This work was achieved in collaboration with Fabien Lotte and Christian Muhl (POTIOC team, Inria-Bordeaux), Moritz Grosse-Wentrup (MPI, Tuebingen), and Reinhold Scherer (TU Graz, Austria).

**A methodological framework for applications combining BCI and VE** Anatole Lécuyer

We have proposed a user-centred methodological framework [46] to guide design and evaluation of applications combining Brain-Computer Interface (BCI) and Virtual Environment (VE). Our framework is based on the contributions of ergonomics to ensure these applications are well suited for end-users. It provides methods, criteria and metrics to perform the phases of the human-centred design process aiming to understand the context of use, specify the user needs and evaluate the solutions in order to define design choices. Several ergonomic methods (e.g., interviews, longitudinal studies, user based testing), objective metrics (e.g., task success, number of errors) and subjective metrics (e.g., mark assigned to an item) are suggested to define and measure the usefulness, usability, acceptability, hedonic qualities, appealingness, emotions related to user experience, immersion and presence to be respected. The benefits and contributions of our user centred framework for the ergonomic design of applications combining BCI and VE were also discussed.

This work was achieved in collaboration with Fabien Lotte from POTIOC team (Inria-Bordeaux).

**HYCOMES Team**

# 6. New Results

## 6.1. Highlights of the Year

The main advances in 2014 of the Hycomes team have been as follows:

**Causality analysis of hybrid systems with ordinary differential equations (ODE)** We have proposed a causality analysis, in the form of a simple type system, rejecting hybrid programs with algebraic circuits — see section 6.2 .

**An index theory of DAE hybrid systems with differential algebraic equations (DAE)** We have proposed a conservative extension of the notion of differentiation index to hybrid systems with differential algebraic equations — see section 6.3 .

## 6.2. A Type-Based Analysis of Causality Loops In Hybrid Systems Modelers

Explicit hybrid systems modelers like Simulink / Stateflow allow for programming both discrete- and continuous-time behaviors with complex interactions between them. A key issue in their compilation is the static detection of algebraic or causality loops. Such loops can cause simulations to deadlock and prevent the generation of statically scheduled code. In [6] (also published as a deliverable of the Sys2Soft collaborative project [14], see 7.2 ), we address this issue for a hybrid modeling language that combines synchronous Lustre-like data-flow equations with Ordinary Differential Equations (ODEs). We introduce the operator last(x) for the left-limit of a signal x. This operator is used to break causality loops and permits a uniform treatment of discrete and continuous state variables. The semantics relies on non-standard analysis, defining an execution as a sequence of infinitesimally small steps. A signal is deemed causally correct when it can be computed sequentially and only progresses by infinitesimal steps outside of discrete events. The causality analysis takes the form of a simple type system. In well-typed programs, signals are proved continuous during integration and can be translated into sequential code for integration with off-the-shelf ODE solvers. The effectiveness of this system is illustrated with several examples written in Zélus, a Lustre-like synchronous language extended with hierarchical automata and ODEs.

## 6.3. On the index of multi-mode DAE Systems

Hybrid systems modelers exhibit a number of difficulties related to the mix of continuous and discrete dynamics and sensitivity to the discretization scheme. Modular modeling, where subsystems models can be simply assembled with no rework, calls for using Differential Algebraic Equations (DAE). In turn, DAE are strictly more difficult than ODE. In most modeling and simulation tools, before simulation can occur, sophisticated pre-processing is applied to DAE systems based on the notion of differentiation index. Graph based algorithms such as the one originally proposed by Pantelides [47] are efficient at finding the differentiation index of a DAE system, structurally (i.e., outside some exceptional values for the system parameters), solving the consistent initialisation problem and, transforming a DAE system into a statically scheduled system of ordinary differential equations (ODE) and implicit functions. The differentiation index for DAE explicitly relies on everything being differentiable. Therefore, extensions to hybrid systems must be done with caution — to our knowledge, no such extension exists, supported by a rigourous mathematical theory. In [8], we use non-standard analysis for this. Non-standard analysis formalizes differential equations as discrete step transition systems with an infinitesimal time basis. This allows to map hybrid DAE systems to difference Algebraic Equations (dAE), for which the notion of difference index can be used. The difference index of a dAE is an easy transposition of the differentiation index of a DAE, where forward shift in time (using a next() operator) replaces differentiation. We prove that the differentiation index of a DAE is structurally equal to the difference index of the dAE resulting from its non-standard interpretation. We can thus propose

the difference index of the non-standard semantics of a hybrid DAE system, as a consistent extension of both the differentiation index of DAE and the difference index of dAE. It turns out that the index theory for (discrete time) dAE systems is interesting in itself and raises new issues. We have investigated graph based method similar to the Pantelides [47] algorithm for computing the difference index of a dAE.

## 6.4. A Unifying View of Loosely Time-Triggered Architectures

Cyber-Physical Systems require distributed architectures to support safety critical real-time control. Hermann Kopetz' Time-Triggered Architecture (TTA) has been proposed as both an architecture and a comprehensive paradigm for systems architecture, for such systems. TTA offers the programmer a logical discrete time compliant with synchronous programming, together with timing bounds. A clock synchronization protocol is required, unless the local clocks used themselves provide the recquired accuracy. To relax the strict requirements on synchronization imposed by TTA, Loosely Time-Triggered Architectures (LTTA) have been proposed. In LTTA, computation and communication units are all triggered by autonomous, unsynchronized, clocks. Communication media act as shared memories between writers and readers and communication is non blocking. This is at the price of communication artifacts (such as duplication or loss of data), which must be compensated for by using some "LTTA protocol". In [7] we have pursued our previous work by providing a unified presentation of the two variants of LTTA (token- and time-based), with simplified analyses. We compared these two variants regarding performance and robustness and we provide ways to combine them.

## I4S Project-Team

# 6. New Results

## 6.1. Highlights of the Year

The team organized the 7th European Workshop on SHM in Nantes in July 2014 (http://ewshm2014.com) .

## 6.2. Analysis and control of systems

### *6.2.1. Optimal vibration damping of large structures*

**Participant:** Dominique Siegert.

This paper deals with the theoretical and experimental analysis of magnetically tuned mass dampers, applied to the vibration damping of large structures of civil engineering interest. Two devices are analysed, for which both the frequency tuning ratio and the damping coefficient can be easily and finely calibrated. They are applied for the damping of the vibrations along two natural modes of a mock-up of a bridge under construction. An original analysis, based on the Maxwell receding image method, is developed for estimating the drag force arising inside the damping devices. It also takes into account self inductance effects, yielding a complex nonlinear dependence of the drag force on the velocity. The analysis highlights the range of velocities for which the drag force can be assumed of viscous type, and shows its dependence on the involved geometrical parameters of the dampers. The model outcomes are then compared to the corresponding experimental calibration curves. A dynamic model of the controlled structure equiped with the two damping devices is presented, and used for the development of original optimization expressions and for determining the corresponding maximum achievable damping. Finally, several experimental results are presented, concerning both the free and harmonically forced vibration damping of the bridge mock-up, and compared to the corresponding theoretical predictions. The experimental results reveal that the maximum theoretical damping performance can be achieved, when both the tuning frequencies and damping coefficients of each device are finely calibrated according to the optimization expressions [13], [44].

### *6.2.2. Particle filtering techniques for monitoring of structures*

**Participant:** Laurent Mevel.

The focus of this paper is Bayesian modal parameter recursive estimation based on an interacting Kalman filter algorithm with decoupled distributions for frequency and damping. Interacting Kalman filter is a combination of two widely used Bayesian estimation methods: the particle filter and the Kalman filter. Some sensitivity analysis techniques are also proposed in order to deduce a recursive estimate of modal parameters from the estimates of the damping/stiffness coefficients [28].

### *6.2.3. Uncertainty quantification*

**Participants:** Michael Doehler, Laurent Mevel.

For applications as Operational Modal Analysis (OMA) of vibrating structures, an output-only LTI system with state and measurement noise can be identified using subspace methods. While these identification techniques have been very suitable for the identification of such mechanical, aeronautical or civil structures, covariance expressions of the estimates of the system matrices are difficult to obtain and theoretical results from literature are hard to implement for output-only systems with unknown noise properties in practice. Moreover, the model order of the underlying system is generally unknown and due to noise and model errors, usual statistical criteria cannot be used. Instead, the system is estimated at multiple model orders and some GUI driven stabilization diagram containing the resulting modal parameters is used by the structural engineer. Then, the covariance of the estimates at these different model orders is an important information for the engineer, which, however, would be computationally expensive to obtain with the existing tools. Recently a fast multi-order version of the stochastic subspace identification approach has been proposed, which is based on the use of the QR decomposition of the observability matrix at the largest model order. In this paper, the corresponding covariance expressions for the system matrix estimates at multiple model orders are derived and successfully applied on real vibration data [36], [38].

### 6.2.4. Periodic systems

**Participants:** Ivan Guéguen, Laurent Mevel.

The modal analysis of a wind turbine has been generally handled with the assumption that this structure can be accurately modeled as linear time-invariant. Such assumption may be misleading for stability analysis, especially, with the current development of very large wind turbines with complex dynamic behavior (nonlinearity, aeroelastic coupling). Therefore in this paper, the inherent periodically time-varying dynamics of wind turbines (and for rotating systems, in general) is taken into account. Recently a subspace algorithm for modal analysis of rotating systems has been proposed. It is tested on a simulated and real data from a wind turbine [20], [41].

### 6.2.5. Identification of finite impulse response systems based on quantized output measurements – a quadratic programming-based method

**Participant:** Qinghua Zhang.

This work has been carried out in collaboration with Jiandong Wang (Peking University, China).

Quantized data are typically produced by the process of analog-to-digital conversion and have been widely studied in signal encoding and digital representation. In system identification, the processed data are usually collected after a quantization procedure, but the effect of quantization is often ignored. The study on system identification based on quantized data makes sense when the data are coded with few quantization levels, to the point that the effect of quantization becomes important. In this work we propose a quadratic programming (QP)-based method for identification of finite impulse response (FIR) dynamic systems from quantized or binary data. The main idea of the proposed method is to reformulate this identification problem, usually viewed as a nonlinear estimation problem with discontinuous nonlinearities, in the form of a standard QP problem, which is a convex optimization problem and can be solved efficiently. The complete input conditions ensuring the strict convexity of the QP problem are developed, and the consistency of the estimated parameters is established under the complete input conditions. The results of this study have been published in [27].

### 6.2.6. Wiener System Identification by Weighted Principal Component Analysis

**Participant:** Qinghua Zhang.

This work has been carried out in collaboration with Vincent Laurain (CRAN/CNRS/Université de Lorraine).

A Wiener system consists of two subsystems connected in series, with a linear dynamic subsystem preceding a static nonlinearity. In the field of control systems, the dynamics of a nonlinear system can often be linearized around its working point. Nevertheless, if its output sensor is affected by strongly nonlinear distortions, the linearization of the sensor characteristics may induce large modeling errors. In such situations, Wiener system model is more appropriate than fully linearized models. Wiener system identification is investigated in this work with a finite impulse response (FIR) model of the linear subsystem. Under the assumption of Gaussian input distribution, this work mainly aims at addressing a deficiency of the well-known correlation-based method for Wiener system identification: it fails when the nonlinearity of the Wiener system is an even function. This method is, in the considered Gaussian input case, equivalent to the best linear approximation (BLA) method, which exhibits the same deficiency. Our new method is based on a weighted principal component analysis (wPCA). Its consistency is proved for Wiener systems with either even or non even nonlinearities. Its computational cost is almost the same as that of a standard PCA.The results of this study have been presented at [51].

### 6.2.7. Industrial process for road buildings

**Participant:** Jean Dumoulin.

The increasing use of the baffled-rotary kiln equipment in many innovative materials processing industrial applications suggests examining the heat transfer phenomena in order to improve the multi-phase flow modeling tools. Their development and use will be relevant for tackling the current energy issues. The heat transfer models available for the rotary kiln in the literature are, for now, not enough efficient for the baffled-rotary kiln case. The present paper is aimed at suggesting a wall heat transfer correlation for the rotary kilns with the secondary inlet. The experimental thermal data acquired within large-scale rotary drum applied to the asphalt concrete materials production, are remained in order to give rise the new issues. These latter results are connected to a visualization campaign performed at the pilot-scale in order to assess the transversal distribution of the granular phase materials. Their analysis suggests a more appropriate physical modelling of the wall heat transfer path. It leads to transform the classical correlation of type $Nu = f(Re, Pr)$ in a new expression of type $Nu = f(Re, St)$ based on a new physical modeling inventory corresponding to the hot and cold fluxes flowing within the baffled-rotary kiln. Thus, the major modification is based on the introduction of the Stanton ($St$) number in the wall heat transfer correlation. This expression is found more convenient for the baffled-rotary kiln application. This new expression is validated by the comparison with the experimental Nusselt numbers calculated from the inner heat transfer measurements coefficient measured in the baffled-rotary kiln performed at large scale [21].

### *6.2.8. Industrial process for concrete structure reparation*

**Participant:** Jean Dumoulin.

In civil engineering, reinforced concrete repair by CFRP is a strengthening technique that has proven successfully in the past. The present study is aimed at using thermoplastic CFRP sheets applied and glued under heat. In this research framework, active thermography is used to accomplish two roles: control of the operating temperature of the thermoplastic CFRP sheets during the installation process and evaluation of the bonding quality after welding. The pap er presents results obtained in laboratory with a dedicated test bench coupled with numerical simulations of the process [49].

### *6.2.9. Building energy management*

**Participants:** Alexandre Nassiopoulos, Jordan Brouns.

Problems such as parameter identification for model calibration, optimal design or optimal energy management can all be formulated in a similar framework as problems consisting in finding the minimum of a cost function. The paper presents the software ReTrofiT that specifically treats this kind of problems applied to building energy performance models. ReTrofiT is first of all a simulation tool for evaluating building thermal behavior and computing energy consumptions. The novelty compared to state-of-the-art energy simulation software is that it also integrates a generic set of tools and algorithms to set up and solve optimization problems related to the building thermal model. The use of the adjoint model, that is intrinsically implemented in the code, constructs fast and efficient algorithms to solve linear, non linear, constrained or unconstrained problems addressing a wide range of applications [43].

## 6.3. damage detection for mechanical structures

### *6.3.1. Damage detection and localisation*

**Participants:** Michael Doehler, Luciano Gallegos, Laurent Mevel.

The Stochastic Dynamic Damage Locating Vector approach is a vibration-based damage localization method based on a finite element model of a structure and output-only measurements in both reference and damaged states. A stress field is computed for loads in the null space of a surrogate of the change in the transfer matrix at the sensor positions for some values in the Laplace domain. Then, the damage location is related to positions where the stress is close to zero. Robustness of the localization information can be achieved by aggregating results at different values in the Laplace domain. So far, this approach and in particular the aggregation is deterministic and does not take the uncertainty in the stress estimates into account. In this paper, the damage localization method is extended with a statistical framework. The uncertainty in the output-only measurements is propagated to the stress estimates at different values of the Laplace variable and these

estimates are aggregated based on statistical principles. The performance of the new statistical approach is demonstrated both in a numerical application and a lab experiment, showing a significant improvement of the robustness of the method due to the statistical evaluation of the localization information [22], [37].

### 6.3.2. An Innovations Approach to Fault Diagnosis in Linear Time-Varying Descriptor Systems

**Participant:** Qinghua Zhang.

This work has been carried out in collaboration with Abdouramane Moussa-Ali (LSIS/CNRS/Université de Toulon).

Many modern engineering systems can be modeled by explicit ordinary differential equations (ODE) in state-space form. Such state-space equations have a long-term mathematical history, and a large number of analytical and numerical tools have been developed for their study. Nevertheless, some systems cannot be described by such explicit state-space models, but described by *implicit* differential equations, known as differential-algebraic equations (DAE). After linearization along a trajectory and discretization in time, a nonlinear DAE system is approximately described by *implicit* discrete time state-space equations, known as *descriptor system equations*. In this work, fault diagnosis is studied for time varying descriptor systems. The Kalman filter for descriptor systems is first revisited by completing existing results about its properties that are essential for the purpose of fault diagnosis. Based on the analysis of the effects of the considered actuator and sensor faults on the innovation of the descriptor system Kalman filter, it is shown that the considered fault diagnosis problem in time varying descriptor systems is equivalent to a classical linear regression problem formulated by appropriately filtering the input-output data. Following this result, algorithms for fault diagnosis through maximum likelihood estimation are then developed. The results of this study have been presented at [42].

### 6.3.3. Statistical detection and isolation of additive faults in linear time-varying systems

**Participant:** Qinghua Zhang.

This work has been carried out in collaboration with Michèle Basseville (IRISA/CNRS).

Model-based approaches to fault detection and isolation (FDI) have been mostly studied in the literature for linear time invariant (LTI) systems. In practice, quite often time-varying and/or nonlinear properties of the monitored system cannot be neglected. One of the possible approachs to dealing with nonlinear systems is based on the linearization along the actual or nominal trajectory of the monitored system. Such a linearization generally leads to linear time-varying (LTV) systems, whereas the more basic LTI approximation is usually related to the linearization around a single working point. It is thus clear that methods for FDI in LTV systems are much more powerful than their LTI counterparts. In the present work, we address the FDI problem for LTV systems subject to parametric additive faults. The proposed approach is statistical, by combining a generalized likelihood ratio (GLR) test with the Kalman filter that cancels out the dynamics of the faults effects in the considered LTV systems. With this approach, it is possible to perform fault isolation when the number of sensors is smaller than the number of assumed faults, under an appropriate assumption about the excitation of the system. The results of this study have been published in [29].

### 6.3.4. Robust subspace damage detection

**Participants:** Michael Doehler, Laurent Mevel.

In the last ten years, monitoring the integrity of the civil infrastructure has been an active research topic, including in connected areas as automatic control. It is common practice to perform damage detection by detecting changes in the modal parameters between a reference state and the current (possibly damaged) state from measured vibration data. Subspace methods enjoy some popularity in structural engineering, where large model orders have to be considered. In the context of detecting changes in the structural properties and the modal parameters linked to them, a subspace-based fault detection residual has been recently proposed and applied successfully, where the estimation of the modal parameters in the possibly damaged state is avoided. However, most works assume that the unmeasured ambient excitation properties during measurements of the structure in the reference and possibly damaged condition stay constant, which is hardly satisfied by any application. This paper addresses the problem of robustness of such fault detection methods. It is explained

why current algorithms from literature fail when the excitation covariance changes and how they can be modified. Then, an efficient and fast subspace-based damage detection test is derived that is robust to changes in the excitation covariance but also to numerical instabilities that can arise easily in the computations. Three numerical applications show the efficiency of the new approach to better detect and separate different levels of damage even using a relatively low sample length [18], [35], [17].

### 6.3.5. Sensor placement

**Participant:** Michael Doehler.

Deciding on the position of sensors by optimizing the utility of the monitoring system over a structure lifetime is typically forbidden by computational cost. Sensor placement strategies are, instead, usually formulated for a pre-selected number of sensors and are based on cost functions that can be evaluated for any arrangement without the need for simulations. This paper examines the performance of two such schemes, the first one is derived directly from a technique that detects damage from the shift of a chi-square distribution from central to non-central and takes the optimal arrangement as the one that maximizes the sensitivity of the non-centrality to all parameter changes of equal norm. The second scheme selects the sensor arrangement as that which maximizes a weighted version of the norm of the sensitivity of the covariance of the output to all feasible changes in system parameters. The performance of the two schemes is tested in simulations [32].

### 6.3.6. Reflectometry for external post-tensioned cable monitoring

**Participant:** Qinghua Zhang.

This work has been carried out in collaboration with IFSTTAR, EDF, ENS Cachan and Andra.

Nowadays a considerable number of bridges is reaching an age when repairs become necessary. In some bridges, external post-tension cables are placed in ducts within which the residual internal space is imperfectly filled with a fluid cement grout. Detecting the defaults of filling is visually impossible from the outside. Among non-destructive detection techniques proposed for cable health monitoring, reflectometry techniques offer remarkable advantages in that they can monitor cables in concrete deviator (embedded in concrete) and they do not require human intervention inside the bridge. In this work, the application of reflectometry techniques to cable health monitoring has been investigated via numerical simulations and laboratory experiments. The results of this study have been presented at [53].

### 6.3.7. Efficient Computation of Minmax Tests for Fault Isolation and Their Application to Structural Damage Localization

**Participants:** Michael Doehler, Laurent Mevel.

Fault detection and isolation can be handled by many different approaches. This paper builds upon a hypothesis test that checks whether the mean of a Gaussian random vector has become non-zero in the faulty state, based on a chi2 test. For fault isolation, it has to be decided which components in the parameter set of the Gaussian vector have changed, which is done by variants of the chi2 hypothesis test using the so-called sensitivity and minmax approaches. While only the sensitivity of the tested parameter component is taken into account in the sensitivity approach, the sensitivities of all parameters are used in the minmax approach, leading to better statistical properties at the expense of an increased computational burden. The computation of the respective test variable in the minmax test is cumbersome and may be ill-conditioned especially for large parameter sets, asking hence for a careful numerical evaluation. Furthermore, the fault isolation procedure requires the repetitive calculation of the test variable for each of the parameter components that are tested for a change, which may be a significant computational burden. In this paper, dealing with the minmax problem, we propose a new efficient computation for the test variables, which is based on a simultaneous QR decomposition for all parameters. Based on this scheme, we propose an efficient test computation for a large parameter set, leading to a decrease in the numerical complexity by one order of magnitude in the total number of parameters. Finally, we show how the minmax test is useful for structural damage localization, where an asymptotically Gaussian residual vector is computed from output-only vibration data of a mechanical or a civil structure [39].

### 6.3.8. Inverse problems in damage detection

**Participant:** Dominique Siegert.

Reinforced concrete beams are widely employed in civil engineering structures. To reduce the maintenance financial cost, structure damages have to be detected early. To this end, one needs robust monitoring techniques. The paper deals with the identification of mechanical parameters, useful for Structural Health Monitoring, in a 2D beam using inverse modeling technique. The optimal control theory is employed. As an example, we aim to identify a reduction of the steel bar cross-section and a decrease of the concrete Young modulus in damaged areas. In our strategy, the beam is instrumented with strain sensors, and a known dynamic load is applied. In the inverse technique, two space discretizations are considered: a fine dicretization to solve the structural dynamic problem and a coarse discretization for the beam parameter identification. To get the beam parameters, we minimize a classical data misfit functional using a gradient-like algorithm. A low-cost computation of the functional gradient is performed using the adjoint equation. The inverse problem is solved in a general way using engineer numerical tools: Python scripts and the free finite element software Code Aster. First results show that a local reduction of the steel bar cross-section and a local decrease of concrete Young modulus can be detected using this inverse technique [25].

### 6.3.9. NDT by active thermography coupled with infrared shearography

**Participant:** Jean Dumoulin.

As infrastructures are aging, the evaluation of their health is becoming crucial. To do so, numerous Non Destructive Testing (NDT) methods are available. Among them, thermal shearography and active infrared thermography represent two full field and contactless methods for surface inspection. The synchronized use of both methods presents multiples advantages. Most importantly, both NDT are based on different material properties. Thermography depend on the thermal properties and shearography on the mechanical properties. The cross-correlation of both methods result in a more accurate and exact detection of the defects. For real site application, the simultaneous use of both methods is simplified due to the fact that the excitation method (thermal) is the same. Active infrared thermography is the measure of the temperature by an infrared camera of a surface subjected to heat flux. Observation of the variation of temperature in function of time reveal the presence of defects. On the other hand, shearography is a measure of out-of-plane surface displacement. This displacement is caused by the application of a strain on the surface which (in our case) take the form of a temperature gradient inducing a thermal stress [56], [47], [48].

## 6.4. Long term monitoring of civil engineering structure

### 6.4.1. ICT based software for thermal field long term monitoring of civil engineering structures

**Participants:** Antoine Crinière, Jean Dumoulin.

Aging 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. A multi-sensing techniques system, able to date and synchronize measurements carried out by infrared thermography coupled with various measurements data (i.e. weather parameters), have been designed, developed and implemented on real site. This smart sensor called IrLaw/SENSORBOX has been upgraded in order to reach full autonomy and its able to monitor over years civil engineering structures [55], [15], [34].

### 6.4.2. Long term structural health monitoring architecture

**Participant:** Jean Dumoulin.

This work gives a brief description of the main activities and outcomes of the Integrated System for Transport Infrastructures surveillance and Monitoring by Electromagnetic Sensing (ISTIMES – www.istimes.eu) project, which was concerned with the development and implementation of a system able to couple the capabilities of long-term monitoring and quick damage assessment of the critical transport infrastructures. This was performed thanks to the integrated use of the novel and state of art concepts of Earth observation, ground-based sensing techniques and ICT architecture [45], [46].

## 6.5. Material characterization

### *6.5.1. Quantitative non destructive testing in civil engineering*

**Participants:** Jordan Brouns, Antoine Crinière, Jean Dumoulin, Alexandre Nassiopoulos.

By the aging of civil engineering structures a crucial need of reparation or reinforcement appeared through years. This can be done using bonded CFRP plate to assure the mechanical behavior of the structure. This type of reparation need diagnosis to insure the reliability of the reparation procedure. This part focus on the development of 1D to 3D method to asses the quantitative non destructive testing of a repaired structure thanks to active thermography (see [14] and [52]).

### *6.5.2. Thermo-physical characterization for civil engineering application*

**Participant:** Jean Dumoulin.

This papers presents the development of a new device for the determination of thermal conductivity and diffusivity of anisotropic composite plates. The excitation signal is provided through a thermoelectric cooler and does not require any optical source like a laser source for instance. Infrared thermography is used to follow apparent surface temperature evolution with time. Experiments were carried out two composite sample systems (with two different fiber orientations). Result analysis is presented and discussed [40].

### *6.5.3. Emissivity characterization for civil engineering applications*

**Participant:** Jean Dumoulin.

The knowledge of the infrared emissivity of materials used in buildings and civil engineering structures is useful for two specific approaches. First, quantitative diagnosis of buildings or civil engineering infrastructures by infrared thermography requires emissivity values in the spectral bandwidth of the camera used for measurements, in order to obtain accurate surface temperatures; for instance, emissivity in the band III domain is required when using cameras with uncooled detectors (such as micro-bolometer arrays). Second, setting up accurate thermal balances by numerical modeling requires the total emissivity value for a large wavelength domain; this is, for instance, the case for computing the road surface temperature to predict ice occurrence. Furthermore, periodical surveys of emissivity variations due to aging or soiling of surfaces could be useful in many situations such as thermal mapping of roads or building insulation diagnosis. The use of portable emissivity measurement devices is required for that purpose. A device using an indirect measurement method was previously developed in our lab; the method uses measurement of the reflectivity from a modulated IR source and requires calibration with a highly reflective surface. However, that device uses a low-frequency, thermal modulation well adapted to laboratory measurements but unfit for fast and in situ measurements. Therefore, a new, portable system which retains the principle of an indirect measurement but uses a faster-frequency, mechanical modulation more appropriate to outdoor measurements was developed. Both devices allow measurements in the broad ($1\mu$m to $40\mu$m) and narrow ($8\mu$m to $40\mu$m) bands. Experiments were performed on a large number of materials commonly used in buildings and civil engineering structures.The final objective of this work is to build a database of emissivity of these materials. A comparison of laboratory and on-site measurements of emissivity values obtained in both spectral bands will be presented along with an estimation and an analysis of measurement uncertainties [23].

## 6.6. Vision under environmental conditions

### *6.6.1. Infrared Imaging under environmental conditions*

**Participant:** Jean Dumoulin.

An infrared system has been developed to monitor transport infrastructures in a standalone configuration. It is based on low cost infrared thermal cameras linked with a calculation unit in order to produce a corrected thermal map of the surveyed structure at a selected time step. With the inline version, the data collected feed simplified radiative models running a GPU. With the offline version, the thermal map can be corrected when data are collected under different atmospheric conditions up to foggy night conditions. A model for radiative transmission prediction is proposed and limitations are addressed. Furthermore, the results obtained by image and signal processing methods with data acquired on the transport infrastructure opened to traffic are presented. Finally, conclusions and perspectives for new implementation and new functionalities are presented and discussed [16].

### 6.6.2. *Long term thermal monitoring by uncooled infrared camera*

**Participant:** Jean Dumoulin.

Being able to perform easily non-invasive diagnostics for surveillance and monitoring of critical transport infrastructures is a major preoccupation of many technical offices. Among all the existing electromagnet-icmethods, long term thermal monitoring by uncooled infrared camera is a promising technique due to its dissemination potential according to its low cost on the market. Nevertheless, Knowledge of environmental parameters during measurement in outdoor applications is required to carry out accurate measurement corrections induced by atmospheric effects at ground level. Particularly considering atmospheric effects and measurements in foggy conditions close as possible to those that can be encountered around transport infrastructures, both in visible and infrared spectra. In the present study, atmospheric effects are first addressed by using data base available in literature and modelling. Atmospheric attenuation by particles depends greatly of aerosols density, but when relative humidity increases, water vapor condenses onto the particulates suspended in the atmosphere. This condensed water increases the size of the aerosols and changes their composition and their effective refractive index. The resulting effect of the aerosols on the absorption and scattering of radiation will correspondingly be modified [54].

### 6.6.3. *Handling of fog conditions by infrared cameras*

**Participant:** Jean Dumoulin.

Fog conditions are the cause of severe car accidents in western countries because of the poor induced visibility. Its forecast and intensity are still very difficult to predict by weather services. Infrared cameras allow to detect and to identify objects in fog while visibility is too low for eye detection. Over the past years, the implementation of cost effective infrared cameras on some vehicles has enabled such detection. On the other hand pattern recognition algorithms based on Canny filters and Hough transformation are a common tool applied to images. Based on these facts, a joint research program between IFSTTAR and Cerema has been developed to study the benefit of infrared images obtained in a fog tunnel during its natural dissipation. Pattern recognition algorithms have been applied, specifically on road signs which shape is usually associated to a specific meaning (circular for a speed limit, triangle for an alert, ...). It has been shown that road signs were detected early enough in images, with respect to images in the visible spectrum, to trigger useful alerts for Advanced Driver Assistance Systems [33].

**IPSO Project-Team**

# 5. New Results

## 5.1. Highlights of the Year

- E. Faou was plenary speaker at the CANUM, Congrès d'analyse numérique, France, June 2014
- E. Faou was invited to give two presentations in the Analysis and applied mathematics seminars, Cambridge, UK, February 2014.

## 5.2. Multi-revolution composition methods for highly oscillatory differential equations

In [22], we introduce a new class of multi-revolution composition methods (MRCM) for the approximation of the $N$th-iterate of a given near-identity map. When applied to the numerical integration of highly oscillatory systems of differential equations, the technique benefits from the properties of standard composition methods: it is intrinsically geometric and well-suited for Hamiltonian or divergence-free equations for instance. We prove error estimates with error constants that are independent of the oscillatory frequency. Numerical experiments, in particular for the nonlinear Schrödinger equation, illustrate the theoretical results, as well as the efficiency and versatility of the methods.

## 5.3. Multiscale schemes for the BGK-Vlasov-Poisson system in the quasi-neutral and fluid limits. Stability analysis and first order schemes

In [51], in collaboration with G. Dimarco (University of Ferrara, Italy) and M.-H. Vignal (University of Toulouse), we deal with the development and the analysis of asymptotic stable and consistent schemes in the joint quasi-neutral and fluid limits for the collisional Vlasov-Poisson system. In these limits, the classical explicit schemes suffer from time step restrictions due to the small plasma period and Knudsen number. To solve this problem, we propose a new scheme stable for choices of time steps independent from the small scales dynamics and with comparable computational cost with respect to standard explicit schemes. In addition, this scheme reduces automatically to consistent discretizations of the underlying asymptotic systems. In this first work on this subject, we propose a first order in time scheme and we perform a relative linear stability analysis to deal with such problems. The framework we propose permits to extend this approach to high order schemes in the next future. We finally show the capability of the method in dealing with small scales through numerical experiments.

## 5.4. Asymptotic preserving scheme for a kinetic model describing incompressible fluids

In [52], in collaboration with M. Lemou (CNRS, Université de Rennes 1) and R. Rao, A. Ruhi, M. Sekhar (Indian Institute of Science, India), the kinetic theory of fluid turbulence modeling developed by Degond and Lemou is considered for further study, analysis and simulation. Starting with the Boltzmann like equation representation for turbulence modeling, a relaxation type collision term is introduced for isotropic turbulence. In order to describe some important turbulence phenomenology, the relaxation time incorporates a dependency on the turbulent microscopic energy and this makes difficult the construction of efficient numerical methods. To investigate this problem, we focus here on a multi-dimensional prototype model and first propose an appropriate change of frame that makes the numerical study simpler. Then, a numerical strategy to tackle the stiff relaxation source term is introduced in the spirit of Asymptotic Preserving Schemes. Numerical tests are performed in a one-dimensional framework on the basis of the developed strategy to confirm its efficiency.

## 5.5. Comparison of numerical solvers for anisotropic diffusion equations arising in plasma physics

In [39], in collaboration G. Latu (IRFM, Cadarache), we performed a comparison of numerical schemes to approximate anisotropic diffusion problems arising in tokamak plasma physics. We focus on the spatial approximation by using finite volume method and on the time discretization. This latter point is delicate since the use of explicit integrators leads to a severe restriction on the time step. Then, implicit and semi-implicit schemes are coupled to finite volumes space discretization and are compared for some classical problems relevant for magnetically confined plasmas. It appears that the semi-implicit approaches (using ARK methods or directional splitting) turn out to be the most efficient on the numerical results, especially when nonlinear problems are studied on refined meshes, using high order methods in space.

## 5.6. Asymptotic-Preserving scheme based on a Finite Volume/Particle-In-Cell coupling for Boltzmann- BGK-like equations in the diffusion scaling

In [38], in collaboration with A. Crestetto (University of Nantes), we are concerned with the numerical simulation of the collisional Vlasov equation in the diffusion limit using particles. To that purpose, we use a micro-macro decomposition technique introduced by Bennoune, Lemou and Mieussens. Whereas a uniform grid was used to approximate both the micro and the macro part of the full distribution function in their article, we use here a particle approximation for the kinetic (micro) part, the fluid (macro) part being always discretized by standard finite volume schemes. There are many advantages in doing so: (i) the so-obtained scheme presents a much less level of noise compared to the standard particle method; (ii) the computational cost of the micro-macro model is reduced in the diffusion limit since a small number of particles is needed for the micro part; (iii) the scheme is asymptotic preserving in the sense that it is consistent with the kinetic equation in the rarefied regime and it degenerates into a uniformly (with respect to the Knudsen number) consistent (and deterministic) approximation of the limiting equation in the diffusion regime.

## 5.7. Hamiltonian splitting for the Vlasov-Maxwell equations

In [23], in collaboration with L. Einkemmer (University of Innsbruck), a new splitting is proposed for solving the Vlasov-Maxwell system. This splitting is based on a decomposition of the Hamiltonian of the Vlasov-Maxwell system and allows for the construction of arbitrary high order methods by composition (independent of the specific deterministic method used for the discretization of the phase space). Moreover, we show that for a spectral method in space this scheme satisfies Poisson's equation without explicitly solving it. Finally, we present some examples in the context of the time evolution of an electromagnetic plasma instability which emphasizes the excellent behavior of the new splitting compared to methods from the literature.

## 5.8. A hybrid transport-diffusion model for radiative transfer in absorbing and scattering media

In [35], in collaboration with M. Roger (University of Lyon), C. Caliot (CNRS) and P. Coelho (Instituto Superior Tecnico of Lisboa), a new multi-scale hybrid transport-diffusion model for radiative transfer calculations is proposed. In this model, the radiative intensity is decomposed into a macroscopic component calculated by the diffusion equation, and a mesoscopic component. The transport equation for the mesoscopic component allows to correct the estimation of the diffusion equation, and then to obtain the solution of the linear radiative transfer equation. In this work, results are presented for stationary and transient radiative transfer cases, in examples which concern solar concentrated and optical tomography applications. The Monte Carlo and the discrete-ordinate methods are used to solve the mesoscopic equation. It is shown that the multi-scale model allows to improve the efficiency of the calculations when the medium is close to the diffusive regime. Moreover, the development of methods for coupling the radiative transfer equation with the diffusion equation becomes easier with this model than with the usual domain decomposition methods.

## 5.9. Charge conserving grid based methods for the Vlasov-Maxwell equations

In [26], in collaboration with P. Navaro (CNRS, Strasbourg) and E. Sonnendrücker (IPP Garching, Germany), In this article we introduce numerical schemes for the Vlasov-Maxwell equations relying on different kinds of grid based Vlasov solvers, as opposite to PIC schemes, that enforce a discrete continuity equation. The idea underlying this schemes relies on a time splitting scheme between configuration space and velocity space for the Vlasov equation and on the computation of the discrete current in a form that is compatible with the discrete Maxwell solver.

## 5.10. Improving conservation properties of a 5D gyrokinetic semi-Lagrangian code

In [32], in collaboration with G. Latu, V. Grandgirard, J. Abiteboul, G. Dif-Pradalier, X. Garbet, P. Ghendrih Y. Sarazin (IRFM, Cadarache), M. Mehrenberger (University of Strasbourg) and E. Sonnendrücker (IPP Garching, Germany), we are concerned with gyrokinetic turbulent simulations, where the knowledge of some stationary states can help reducing numerical artifacts. Considering long-term simulations, the qualities of the Vlasov solver and of the radial boundary conditions have an impact on the conservation properties. In order to improve mass and energy conservation mainly, the following methods are investigated: fix the radial boundary conditions on a stationary state, use a 4D advection operator that avoids a directional splitting, interpolate with a delta-f approach. The combination of these techniques in the semi-Lagrangian code gysela leads to a net improvement of the conservation properties in 5D simulations.

## 5.11. Simulations of Kinetic Electrostatic Electron Nonlinear (KEEN) Waves with Variable Velocity Resolution Grids and High-Order Time-Splitting

In [16], in collaboration with B. Afeyan (Polymath Research, USA), F. Casa (University Jaume, Spain), A. Dodhy, E. Sonnendrücker (IPP Garching, Germany) and M. Mehrenberger (University of Strasbourg), we are concerned with KEEN waves which are non-stationary, nonlinear, self-organized asymptotic states in Vlasov plasmas. They lie outside the precepts of linear theory or perturbative analysis, unlike electron plasma waves or ion acoustic waves. Steady state, nonlinear constructs such as BGK modes also do not apply. The range in velocity that is strongly perturbed by KEEN waves depends on the amplitude and duration of the ponderomotive force generated by two crossing laser beams, for instance, used to drive them. Smaller amplitude drives manage to devolve into multiple highly-localized vorticlets, after the drive is turned off, and may eventually succeed to coalesce into KEEN waves. Fragmentation once the drive stops, and potential eventual remerger, is a hallmark of the weakly driven cases. A fully formed (more strongly driven) KEEN wave has one dominant vortical core. But it also involves fine scale complex dynamics due to shedding and merging of smaller vortical structures with the main one. Shedding and merging of vorticlets are involved in either case, but at different rates and with different relative importance. The narrow velocity range in which one must maintain sufficient resolution in the weakly driven cases, challenges

fixed velocity grid numerical schemes. What is needed is the capability of resolving locally in velocity while maintaining a coarse grid outside the highly perturbed region of phase space. We here report on a new Semi-Lagrangian Vlasov-Poisson solver based on conservative non-uniform cubic splines in velocity that tackles this problem head on. An additional feature of our approach is the use of a new high-order time-splitting scheme which allows much longer simulations per computational effort. This is needed for low amplitude runs. There, global coherent structures take a long time to set up, such as KEEN waves, if they do so at all. The new code's performance is compared to uniform grid simulations and the advantages are quantified. The birth pains associated with weakly driven KEEN waves are captured in these simulations. Canonical KEEN waves with ample drive are also treated using these advanced techniques. They will allow the efficient simulation of KEEN waves in multiple dimensions, which will be tackled next, as well as generalizations to Vlasov-Maxwell codes. These are essential for pursuing the impact of KEEN waves in high energy density plasmas and in inertial confinement fusion applications. More generally, one needs a fully-adaptive grid-in-phase-space method which could handle all small vorticlet dynamics whether pealing off or remerging. Such

fully adaptive grids would have to be computed sparsely in order to be viable. This two-velocity grid method is a concrete and fruitful step in that direction.

## 5.12. Gyroaverage operator on polar mesh

In [36], in collaboration with C. Steiner, M. Mehrenberger (University of Strasbourg) V. Grandgirard, G. Latu (IRFM, Cadarache). In this work, we are concerned with numerical approximation of the gyroaverage operators arising in plasma physics to take into account the effects of the finite Larmor radius corrections. The work initiated in [Crouseilles, Mehrenberger, Sellama, CiCP 2010] is extended here to polar geometries. A direct method is proposed in the space configuration which consists in integrating on the gyrocircles using interpolation operator (Hermite or cubic splines). Numerical comparisons with a standard method based on a Pade approximation are performed: (i) with analytical solutions, (ii) considering the 4D drift-kinetic model with one Larmor radius and (iii) on the classical linear DIII-D benchmark case [6]. In particular, we show that in the context of a drift-kinetic simulation, the proposed method has similar computational cost as the standard method and its precision is independent of the radius.

## 5.13. A new fully two-dimensional conservative semi-Lagrangian method: applications on polar grids, from diocotron instability to ITG turbulence

In [25], in collaboration with P. Glanc, S. Hirstoaga, E. Madaule, M. Mehrenberger, J. Pétri (University of Strasbourg), While developing a new semi-Lagrangian solver, the gap between a linear Landau run in 1dx1d and a 5D gyrokinetic simulation in toroidal geometry is quite huge. Intermediate test cases are welcome for checking the code. We consider here as building block, a 2D guiding-center type equation on an annulus. We first revisit a 2D test case previously done with a PIC approach and detail the boundary conditions. We then consider a 4D drift-kinetic slab simulation for which we give some first results of a new conservative method.

## 5.14. Uniformly accurate numerical schemes for highly oscillatory Klein-Gordon and nonlinear Schrödinger equations

In [21], we are interested in the numerical simulation of nonlinear Schrödinger and Klein-Gordon equations. We present a general strategy to construct numerical schemes which are uniformly accurate with respect to the oscillation frequency. This is a stronger feature than the usual so called ”Asymptotic preserving" property, the last being also satisfied by our scheme in the highly oscillatory limit. Our strategy enables to simulate the oscillatory problem without using any mesh or time step refinement, and the orders of our schemes are preserved uniformly in all regimes. In other words, since our numerical method is not based on the derivation and the simulation of asymptotic models, it works in the regime where the solution does not oscillate rapidly, in the highly oscillatory limit regime, and in the intermediate regime with the same order of accuracy. The method is based on two main ingredients. First, we embed our problem in a suitable ”two-scale" reformulation with the introduction of an additional variable. Then a link is made with classical strategies based on Chapman-Enskog expansions in kinetic theory despite the dispersive context of the targeted equations, allowing to separate the fast time scale from the slow one. Uniformly accurate (UA) schemes are eventually derived from this new formulation and their properties and performances are assessed both theoretically and numerically.

## 5.15. Asymptotic preserving schemes for the Wigner-Poisson-BGK equations in the diffusion limit

In [24], we focus on the numerical simulation of the Wigner-Poisson-BGK equation in the diffusion asymptotics. Our strategy is based on a ”micro-macro" decomposition, which leads to a system of equations that couple the macroscopic evolution (diffusion) to a microscopic kinetic contribution for the fluctuations. A semi-implicit discretization provides a numerical scheme which is stable with respect to the small parameter $\varepsilon$ (mean free path) and which possesses the following properties: (i) it enjoys the asymptotic preserving property in the diffusive limit; (ii) it recovers a standard discretization of the Wigner-Poisson equation in the collisionless regime. Numerical experiments confirm the good behaviour of the numerical scheme in both regimes. The case of a spatially dependent $\varepsilon(x)$ is also investigated.

## 5.16. Models of dark matter halos based on statistical mechanics: II. The fermionic King model

In [49] we study the fermionic King model which may provide a relevant model of dark matter halos. The exclusion constraint can be due to quantum mechanics (for fermions such as massive neutrinos) or to Lynden- Bells statistics (for collisionless systems undergoing violent relaxation). This model has a finite mass. Dwarf and intermediate size halos are degenerate quantum objects stabilized against gravitational collapse by the Pauli exclusion principle. Large halos at sufficiently high energies are in a gaseous phase where quantum effects are negligible. They are stabilized by thermal motion. Below a critical energy Ec they undergo gravitational collapse (gravothermal catastrophe). This may lead to the formation of a central black hole that does not affect the structure of the halo. This may also lead to the formation of a compact degenerate object surrounded by a hot massive atmosphere extending at large distances. We argue that large dark matter halos should not contain a degenerate nucleus (fermion ball) because these nucleus-halo structures are thermodynamically unstable. We compare the rotation curves of the classical King model to observations of large dark matter halos (Burkert profile). Because of collisions and evaporation, the central density increases while the slope of the halo density profile decreases until an in- stability takes place. We find that the observations are compatible with a King profile at, or close to, the point of marginal stability in the micro- canonical ensemble. At that point, the King profile can be fitted by the modified Hubble profile. This is qualitatively similar to the Burkert profile and discrepancies between the King model and the observations are interpreted as a result of incomplete relaxation.

## 5.17. Models of dark matter halos based on statistical mechanics: I. The classical King model

In [48] we consider the possibility that dark matter halos are described by the Fermi-Dirac distribution at finite temperature. This is the case if dark matter is a self-gravitating quantum gas made of massive neutrinos at statistical equilibrium. This is also the case if dark matter can be treated as a self-gravitating collisionless gas experiencing Lynden-Bell?s type of violent relaxation. In order to avoid the infinite mass problem and carry out a rigorous stability analysis, we consider the fermionic King model. In this paper, we study the non-degenerate limit leading to the classical King model. This model was initially introduced to describe globular clusters and we propose to apply it also to large dark matter halos where quantum effects are negligible. We study the thermodynamical stability of the different configurations and compare the prediction of the classical King model to the observations of large dark matter halos. Because of collisions and evaporation, the central density increases while the slope of the halo density profile decreases until an instability takes place. We show that large dark matter halos are relatively well-described by the King model at, or close to, the point of marginal microcanonical stability. At that point, the King model generates a density profile that can be approximated by the modified Hubble profile. This profile has a flat core and decreases as r?3 at large distances, like the observational Burkert profile. For large halos, the flat core is due to finite temperature effects, not to quantum mechanics. We argue that statistical mechanics may provide a good description of dark matter halos and interpret the discrepancies as a result of incomplete relaxation like in the case of stellar systems.

## 5.18. Analysis of models for quantum transport of electrons in graphene layers

In [28], two mathematical models for the self consistent quantum transport of electrons in a graphene layer are presented are analyzed. We treat two situations. First, when the particles can move in all the plane $R^2$, the model takes the form of a system of massless Dirac equations coupled together by a selfconsistent potential, which is the trace in the plane of the graphene of the 3D Poisson potential associated to surface densities. Second, we consider a situation where the particles are constrained in a regular bounded domain $\Omega$. In order to take into account Dirichlet boundary conditions which are not compatible with the Dirac Hamiltonian $H_0$, we propose a different model built on a modified Hamiltonian displaying the same energy band diagram as $H_0$ near the Dirac points.

## 5.19. Dimension reduction for anisotropic Bose-Einstein condensates in the strong interaction regime

The work [44] deals with the problem of dimension reduction for the three dimensional Gross-Pitaevskii equation (GPE) describing a Bose-Einstein condensate confined in a strongly anisotropic harmonic trap. Since the gas is assumed to be in a strong interaction regime, we have to analyze two combined singular limits: a semi-classical limit in the transport direction and the strong partial confinement limit in the transversal direction.

## 5.20. Superconvergence of Strang splitting for NLS in $T^d$

In [47], we investigate the convergence properties of semi-discretized approximations by Strang splitting method applied to fast-oscillating nonlinear Schrödinger equations. Our main contribution is to show that Strang splitting with constant step-sizes is unexpectedly more accurate by a factor $\varepsilon$ as compared to established results when the step-size is chosen as an integer fraction of the period, owing to an averaging effect.

## 5.21. Strong confinement limit for the nonlinear Schrödinger equation constrained on a curve

The preprint [58] is devoted to the cubic nonlinear Schrödinger equation in a two dimensional waveguide with shrinking cross section of order $\varepsilon$. For a Cauchy data living essentially on the first mode of the transverse Laplacian, we provide a tensorial approximation of the solution $\psi^\varepsilon$ in the limit $\varepsilon \to 0$, with an estimate of the approximation error, and derive a limiting nonlinear Schrödinger equation in dimension one with an additional effective potential depending on the curvature.

## 5.22. The fermionic King model

In [50], we study the fermionic King model which may provide a relevant model of dark matter halos.

## 5.23. Landau damping in Sobolev spaces for the Vlasov-HMF model

In [56], we consider the Vlasov-HMF (Hamiltonian Mean-Field) model. We consider solutions starting in a small Sobolev neighborhood of a spatially homogeneous state satisfying a linearized stability criterion (Penrose criterion). We prove that these solutions exhibit a scattering behavior to a modified state, which implies a nonlinear Landau damping effect with polynomial rate of damping.

## 5.24. Collisions of vortex filament pairs

In [18], we consider the problem of collisions of vortex filaments for a model introduced by Klein, Majda and Damodaran, and Zakharov to describe the interaction of almost parallel vortex filaments in three-dimensional fluids. Since the results of Crow examples of collisions are searched as perturbations of antiparallel translating pairs of filaments, with initial perturbations related to the unstable mode of the linearized problem; most results are numerical calculations. In this article we first consider a related model for the evolution of pairs of filaments and we display another type of initial perturbation leading to collision in finite time. Moreover we give numerical evidence that it also leads to collision through the initial model. We finally study the self-similar solutions of the model.

## 5.25. Asymptotic preserving schemes for the Klein-Gordon equation in the non-relativistic limit regime

In [30], we consider the Klein-Gordon equation in the non-relativistic limit regime, i.e. the speed of light c tending to infinity. We construct an asymptotic expansion for the solution with respect to the small parameter depending on the inverse of the square of the speed of light. As the first terms of this asymptotic can easily be simulated our approach allows us to construct numerical algorithms that are robust with respect to the large parameter c producing high oscillations in the exact solution.

## 5.26. Analysis of a large number of Markov chains competing for transitions

In [17], we consider the behavior of a stochastic system composed of several identically distributed, but non independent, discrete-time absorbing Markov chains competing at each instant for a transition. The competition consists in determining at each instant, using a given probability distribution, the only Markov chain allowed to make a transition. We analyze the first time at which one of the Markov chains reaches its absorbing state. When the number of Markov chains goes to infinity, we analyze the asymptotic behavior of the system for an arbitrary probability mass function governing the competition. We give conditions for the existence of the asymptotic distribution and we show how these results apply to cluster-based distributed systems when the competition between the Markov chains is handled by using a geometric distribution.

## 5.27. Coexistence phenomena and global bifurcation structure in a chemostat-like model with species-dependent diffusion rates

In [20], we study the competition of two species for a single resource in a chemostat. In the simplest space-homogeneous situation, it is known that only one species survives, namely the best competitor. In order to exhibit coexistence phenomena, where the two competitors are able to survive, we consider a space dependent situation: we assume that the two species and the resource follow a diffusion process in space, on top of the competition process. Besides, and in order to consider the most general case, we assume each population is associated with a distinct diffusion constant. This is a key difficulty in our analysis: the specific (and classical) case where all diffusion constants are equal, leads to a particular conservation law, which in turn allows to eliminate the resource in the equations, a fact that considerably simplifies the analysis and the qualitative phenomena. Using the global bifurcation theory, we prove that the underlying 2-species, stationary, diffusive, chemostat-like model, does possess coexistence solutions, where both species survive. On top of that, we identify the domain, in the space of the relevant bifurcation parameters, for which the system does have coexistence solutions.

## 5.28. Global behavior of N competing species with strong diffusion: diffusion leads to exclusion

In [46], we study the following problem. For a large class of models involving several species competing for a single resource in a *homogeneous* environment, it is known that the competitive exclusion principle holds: only one species survives eventually. Various works indicate though that coexistence of many species is possible when the competition occurs in a *heterogeneous* environment. We propose here a spatially heterogeneous system modeling several species competing for a single resource, and migrating in the spatial domain. For this model, it is known, at least in particular cases, that if migrations are *slow* enough, then coexistence occurs. In this paper we show at variance that if the spatial migrations are *fast* enough, then our system can be approximated by a spatially homogeneous system, called aggregated model, which can be explicitly computed, and we show that if the competitive exclusion principle holds for the aggregated model, then it holds as well for the original, spatially heterogeneous model. In other words, we show the persistence of the competitive exclusion principle in the spatially heterogeneous situation when migrations are fast. As a consequence, for fast migrations only one species may survive, namely the best competitor *in average*. We last study which is the best competitor *in average* on some examples, and draw some ecological consequences.

## 5.29. Randomized Message-Passing Test-and-Set

In [42] we present a solution to the well-known Test&Set operation in an asynchronous system prone to process crashes. Test&Set is a synchronization operation that, when invoked by a set of processes, returns yes to a unique process and returns no to all the others. Recently many advances in implementing Test&Set objects have been achieved, however all of them target the shared memory model. In this paper we propose an implementation of a Test&Set object in the message passing model. This implementation can be invoked by any number $p \leq n$ of processes where $n$ is the total number of processes in the system. It has an

expected individual step complexity in O(log p) against an oblivious adversary, and an expected individual message complexity in O(n). The proposed Test&Set object is built atop a new basic building block, called selector, that allows to select a winning group among two groups of processes. We propose a message-passing implementation of the selector whose step complexity is constant. We are not aware of any other implementation of the Test&Set operation in the message passing model.

## 5.30. Existence of densities for the 3D Navier–Stokes equations driven by Gaussian noise

In [27], we prove three results on the existence of densities for the laws of finite dimensional functionals of the solutions of the stochastic Navier-Stokes equations in dimension 3. In particular, under very mild assumptions on the noise, we prove that finite dimensional projections of the solutions have densities with respect to the Lebesgue measure which have some smoothness when measured in a Besov space. This is proved thanks to a new argument inspired by an idea introduced by N. Fournier and J. Printems.

## 5.31. Diffusion limit for the radiative transfer equation perturbed by a Markovian process

In [54], we study the stochastic diffusive limit of a kinetic radiative transfer equation, which is non linear, involving a small parameter and perturbed by a smooth random term. Under an appropriate scaling for the small parameter, using a generalization of the perturbed test-functions method, we show the convergence in law to a stochastic non linear fluid limit.

## 5.32. Diffusion limit for the radiative transfer equation perturbed by a Wiener process

In [55], we consider the rigorous derivation of a stochastic non-linear diffusion equation from a radiative transfer equation perturbed with a random noise of white noise type. The proof of the convergence relies on a formal Hilbert expansion and the estimation of the remainder. The Hilbert expansion has to be done up to order 3 to overcome some difficulties caused by the random noise.

**KerData Project-Team**

# 6. New Results

## 6.1. Highlights of the Year

IEEE Cluster 2014. The KerData Team had a leading role the organization of the IEEE Cluster 2014 conference, held in Madrid (22–26 September 2014): Gabriel Antoniu as PC Chair, Luc Bougé as Student Mentoring Program Chair, Alexandru Costan as Submissions Chair.

## 6.2. Data Management for Geographically Distributed Workflows

### *6.2.1. OverFlow: a multi-site-aware framework for Big Data management*

**Participants:** Radu Tudoran, Alexandru Costan, Gabriel Antoniu.

The global deployment of cloud datacenters is enabling large-scale scientific workflows to improve performance and deliver fast responses. This unprecedented geographical distribution of the computation coincides with an increase in the scale of the data handled by such applications, bringing new challenges related to the efficient data management across sites. High throughput, low latencies or cost-related trade-offs are just a few concerns for both cloud providers and users when it comes to handling data across datacenters, as shown in earlier evaluations [21]. Existing solutions are limited to cloud-provided storage, which offers low performance based on rigid cost schemes. In turn, workflow engines need to find ad-hoc substitutes, achieving performance at the cost of complex system configurations, maintenance overheads, reduced reliability and reusability.

We tackle these problems by trying to understand to what extent the intra- and inter-datacenter transfers can impact the total makespan of cloud workflows. We advocate storing data on the compute nodes and transferring files between them directly, in order to exploit data locality and to avoid the overhead of interacting with a shared file system. Under these circumstances, we propose a file management service that enables high throughput through self-adaptive selection among multiple transfer strategies (e.g. FTP-based, BitTorrent-based, etc.). Next, we focus on the more general case of large-scale data dissemination across geographically distributed sites. The key idea is to predict I/O and transfer performance accurately and robustly in a dynamic cloud environment in order to decide judiciously how to perform transfer optimizations over federated datacenters: predict the best combination of protocol and transfer parameters (e.g., multi-routes, flow count, multicast enhancement, replication degree) to maximize throughput or minimize costs, according to users policies. We have implemented these principles in OverFlow, as part of the Azure Cloud so that applications could use it using a Software-as-a-Service (SaaS) approach.

OverFlow [20] was validated on the Microsoft cloud across the 6 EU and US sites. The experiments were conducted on hundreds of nodes using synthetic benchmarks and real-life bio-informatics applications (A-Brain, BLAST). The results show that our system is able to model the cloud performance accurately and to leverage this for efficient data dissemination, being able to reduce the monetary costs and transfer time by up to 3 times.

### *6.2.2. Metadata management for geographically distributed workflows*

**Participants:** Luis Eduardo Pineda Morales, Radu Tudoran, Alexandru Costan, Gabriel Antoniu.

Scientific workflow data can reach sizes that exceed single-site capabilities. It is needed to support fine-grain data stripping to handle either very large files or very large sets of small files across data centers. Therefore, metadata becomes a critical issue. Moreover, workflow metadata provides crucial information to optimize data management, particularly in the context of geographically distributed data centers. Many present-day distributed file systems, such as GoogleFS and HDFS, include a potential bottleneck as the number of files grows, because they use a centralized metadata management scheme. Thus, we argue for a new, *cloud-based, distributed metadata management* scheme.

We have designed four different approaches to a geographically distributed metadata registry, namely: a) baseline centralized version; b) distributed on each data center with centralized replication agent; c) decentralized non-replicated; and d) decentralized replicated with hierarchical access. A comparative analysis showed that the later strategy performs best in terms of metadata operations per time unit. We then evaluate each of our approaches against various workflow benchmarks, with the purpose of dynamically adapt the metadata handling scheme according to the underlying application and cloud contexts. In the next phase, we will provide a uniform metadata handling tool for scientific workflow engines across cloud datacenters, as well as derive a cost model to offer users the best trade-off (performance vs. cost) driven by their constraints.

### *6.2.3. Transfer-as-a-Service: a cost-effective model for multi-site cloud data management*

**Participants:** Radu Tudoran, Alexandru Costan, Gabriel Antoniu.

Existing cloud data management solutions are limited to cloud-provided storage, which offers low performance based on rigid cost schemas. Users are therefore forced to design and deploy custom solutions, achieving performance at the cost of complex system configurations, maintenance overheads, reduced reliability and reusability. In [19] we have proposed a dedicated cloud data-transfer service that supports largescale data dissemination across geographically distributed sites, advocating for a Transfer-as-a-Service (TaaS) paradigm. The system aggregates the available bandwidth by enabling multi-route transfers across cloud sites, based on the approach previously described.

We argue that the adoption of such a TaaS approach brings several benefits for both users and the cloud providers who propose it. For users of multi-site or federated clouds, our proposal is able to decrease the variability of transfers and increase the throughput up to three times compared to baseline user options, while benefiting from the well-known high availability of cloud-provided services. For cloud providers, such a service can decrease the energy consumption within a datacenter down to half compared to user-based transfers. Finally, we propose a dynamic cost model schema for the service usage, which enables the cloud providers to regulate and encourage data exchanges via a data transfer market.

## 6.3. Optimizing Map-Reduce processing

### *6.3.1. Optimizing Map-Reduce in virtualized environments*

**Participant:** Shadi Ibrahim.

As data-intensive applications become popular in the cloud, their performance on the virtualized platform calls for empirical evaluations and technical innovations. Virtualization has become a prominent tool in data centers and is extensively leveraged in cloud environments: it enables multiple virtual machines (VMs) — with multiple operating systems and applications — to run within a physical server. However, virtualization introduces the challenging issue of providing effective QoS to VMs and preserving the high disk utilization (i.e., reducing the seek delay and rotation overhead) when allocating disk resources to VMs.

In [32], we developed a novel disk I/O scheduling framework, named *Pregather*, to improve disk I/O efficiency through exposure and exploitation of the spatial locality in the virtualized environment (regional and sub-regional spatial locality corresponds to the virtual disk space and applications' access patterns, respectively). In [14], we extend *Pregather* to improve disk I/O utilization further while reducing the disk resource contention and ensuring the I/O performance of VMs with different degrees of spatial locality. To do so, we developed an adaptive time-slice allocation scheme based on the spatial locality of VMs, to adjust the lengths of I/O time slices of VMs dynamically. We evaluated *Pregather* through extensive experiments that involve multiple simultaneous applications of both synthetic benchmarks and a Map-Reduce application (e.g., distributed sort) on Xen-based platforms.

Our evaluations use synthetic benchmarks, a Map-Reduce application (distributed sort) and database workloads. They demonstrate that *Pregather* achieves high disk spatial locality, yields a significant improvement in disk throughput, ensures the performance guarantees of VMs, and enables improved Hadoop performance. This work was done in collaboration with Hai Jin, Song Wu and Xiao Ling from Huazhong University of Science and Technology (HUST).

### 6.3.2. A simulation approach to evaluate Map-Reduce performance under failure

**Participants:** Tien Dat Phan, Shadi Ibrahim, Gabriel Antoniu, Luc Bougé.

Map-Reduce is emerging as a prominent tool for large-scale data analysis. It is often advocated as an easier-to-use, efficient and reliable replacement for the traditional programming model of moving the data to the computation. The popular open source implementation of Map-Reduce, Hadoop, is now widely used by major companies, including Facebook, Amazon, Last.fm, and the New York Times. Fault tolerance is one of the key features of the Map-Reduce system. Map-Reduce is designed to handle various kind of failures including stop-fail and time failures: Map-Reduce re-executes failed tasks and re-launches another copy of slow tasks. Although many studies have been dedicated to investigate and improve the performance of Map-Reduce, comparatively little attention has been devoted on investigating the performance of Map-Reduce under failures.

In this ongoing work, we investigate how Map-Reduce (i.e., Hadoop) behaves under failures. To do so, we developed *iHadoop*, a Hadoop simulator developed in Java on top of SimGrid. Experimental results demonstrated that *iHadoop* accurately simulates the behavior of Hadoop and therefore can accurately predict the performance of Hadoop when running on large-scale system using the Grid'5000 testbed. In particular, iHadoop can accurately predict the percentage of Map tasks locality, the number of speculative tasks and, more importantly, the overall execution time of Map-Reduce applications under failures.

### 6.3.3. Waste-Free Preemption Strategy for Hadoop

**Participants:** Orçun Yildiz, Shadi Ibrahim, Gabriel Antoniu.

Hadoop is widely used in the computer industry because of its scalability, reliability, ease of use, and low cost of implementation. Hadoop hides the complexity of discovery and handling failures from the schedulers, but the burden of failure recovery relies entirely on users, regardless of root causes. We systematically assess this burden through a set of experiments, and argue that more effort to reduce this cost to users is desirable. We also analyze the drawback of current Hadoop mechanism in prioritizing failed tasks. By trying to launch failed tasks as soon as possible regardless of locality, it significantly increases the execution time of jobs with failed tasks, due to two reasons: 1) available slots might not be free up as quickly as expected; and 2) the slots might belong to machines with no data on it, introducing extra cost for data transfer through network, which is normally the most scare resource in nowadays data centers.

In this ongoing work, we introduce a new algorithmic approach called the waste-free preemption. The waste-free preemption saves Hadoop scheduler from solely choosing between kill, which instantly releases the slots but is wasteful, and wait, which does not waste any previous effort but fails for the two above-mentioned reasons. With this new strategy, a preemptive version of Hadoop's default schedulers (FIFO and Fair) has been implemented. The evaluation demonstrates the effectiveness of the new feature by comparing its performance with the traditional Hadoop mechanism.

### 6.3.4. Optimizing incremental Map-Reduce computations for on-demand data upload

**Participants:** Stefan Ene, Alexandru Costan, Gabriel Antoniu.

Research on cloud-based Big Data analytics has focused so far on optimizing the performance and cost-effectiveness of the computations, while largely neglecting an important aspect: users need to upload massive datasets on clouds for their computations. In this context, we study the problem of running Map-Reduce applications by considering the simultaneous optimization of performance and cost of both the data upload and its corresponding computation taken together. We analyze the feasibility of incremental Map-Reduce approaches to let the computation progress as much as possible during the data upload by using already transferred data to compute intermediate results.

Current approaches that are either optimized for different purposes, or address the computational problem independent of the data upload. In contrast, to our best knowledge, this is the first approach which simultaneously focuses on both data upload and processing. In this context, we show in [17] that it is not always efficient to attempt to overlap the transfer time with as many incremental computations as possible: a better solution is to wait long enough to fill the computational capacity of the Map-Reduce cluster. Based on this idea, we

developed and evaluated a preliminary prototype. To demonstrate the viability of our prototype in real-life, we run extensive experiments in a distributed setting that involves a 11-node large incremental Map-Reduce deployment based on Hourglass. The results show significant benefits for our approach compared with a simple incremental strategy that starts the next incremental job immediately after the previous has finished: the time-to-solution is improved by 1%, the compute time after the data transfer is finished is reduced by up to 40% and the cost is reduced 10 %-44 %. Compared with a serialized strategy that starts the computation only after all data is transferred, the time-to-solution is improved by up to 30 %, the compute time after the upload finished is reduced by up to 60 % and the cost is reduced between 4 % and 23 %.

## 6.4. Energy-Aware Data Management in the Cloud and Exascale HPC Systems

### *6.4.1. Energy-efficiency in Hadoop*

**Participants:** Tien Dat Phan, Shadi Ibrahim, Gabriel Antoniu, Luc Bougé.

With increasingly inexpensive cloud storage and increasingly powerful cloud processing, the cloud has rapidly become the environment to store and analyze data. Most of the large-scale data computations in the cloud heavily rely on the Map-Reduce paradigm and its Hadoop implementation. Nevertheless, this exponential growth in popularity has significantly impacted power consumption in cloud infrastructures.

In [18], we focus on Map-Reduce and we investigate the impact of dynamically scaling the frequency of compute nodes on the performance and energy consumption of a Hadoop cluster. To this end, a series of experiments are conducted to explore the implications of Dynamic Voltage Frequency scaling (DVFS) settings on power consumption in Hadoop-clusters. By adapting existing DVFS governors (i.e., *performance*, *power-save*, *on-demand*, *conservative* and *user-space*) in the Hadoop cluster, we observe significant variation in performance and power consumption of the cluster with different applications when applying these governors: the different DVFS settings are only sub-optimal for different Map-Reduce applications. Furthermore, our results reveal that the current CPU governors do not exactly reflect their design goal and may even become ineffective to manage power consumption in Hadoop clusters.

More recently, we extended our work to further illustrate the behavior of different governors, which influence the energy consumption in Hadoop Map-Reduce. We extend our experimental platform from 15 to 40 nodes and we employ two additional benchmarks: K-means and wordcount. Moreover, we investigate preliminary DVFS models that adjust to the various stages of Hadoop applications. We also demonstrate that achieving better energy efficiency in Hadoop cannot be done by tuning the governors parameters, nor through a naive coarse-grained tuning of the CPU frequencies or the governors according the running phase (i.e., map phase or reduce phase). In addition, we provide an extensive discussion of the sensitivity for different parameters employed in *ondemand* and *conservative* governors.

### *6.4.2. Exploring the impact of dedicated resources on energy consumption in Exascale systems*

**Participants:** Orçun Yildiz, Matthieu Dorier, Shadi Ibrahim, Gabriel Antoniu.

The advent of fast, unprecedentedly scalable, yet energy-hungry Exascale supercomputers poses a major challenge consisting in sustaining a high performance-per-Watt ratio. While much recent work has explored new approaches to I/O management, aiming to reduce the I/O performance bottleneck exhibited by HPC applications (and hence to improve application performance), there is comparatively little work investigating the impact of I/O management approaches on energy consumption.

In [23], we explore how much energy a supercomputer consumes while running scientific simulations when adopting various I/O management approaches. We closely examine three radically different I/O schemes including time partitioning, dedicated cores, and dedicated nodes. We implement the three approaches within the Damaris I/O middleware and perform extensive experiments with one of the target HPC applications of the Blue Waters sustained-Petaflops supercomputer project: the CM1 atmospheric model. The experimental results obtained on the French Grid'5000 platform highlight the differences between these three approaches and illustrate in which way various configurations of the application and of the system can impact performance and energy consumption.

Based on those experimental results, we are working on building a new energy model which can estimate the energy consumptions of various I/O management approaches and help users in selecting the optimal I/O approach to run their application.

### 6.4.3. Energy impact of data consistency management in the HBase distributed cloud data store

**Participants:** Álvaro García Recuero, Shadi Ibrahim, Gabriel Antoniu.

Cloud Computing has recently emerged as a key technology providing individuals and companies with access to remote computing and storage infrastructures. In order to achieve high-availability and fault-tolerance, cloud data storage relies on replication. That comes with the issue of consistency among distant replicas so one can always get the most up-to-date values from any of them (*e.g.*, fresh data).

In that context, being able to provide data consistency and continuous availability in the Cloud is yet a non-trivial problem, mainly due to the ever-increasing volume, variety and velocity of data in storage systems. Big data processing engines (e.g., Hadoop, Spark, etc.) as well as modern NoSQL storage back-ends (HBase, Cassandra) have to therefore deal with these high volumes of information at large scale while still providing applications with a consistent and on-time data delivery.

In this work, a set of synthetic workloads from YCSB (Yahoo! Cloud Service Benchmark) was configured to simulate random reads/writes and measure their impact into the overall energy consumption of a well-known distributed data store, HBase. The cluster is comprised of 40 servers and the results have been confirmed with several configurations and runs on the Grid5000 experimental platform. The results indicate that certain write-intensive workloads can be a bottleneck in terms of throughput, further deepening the problem of having an energy-efficient consistency management. Regarding read-intensive workloads, we observe similar patterns but with a very different impact on their energy footprint. We plan to further investigate how to leverage energy-aware mechanisms that overcome the energy-consistency trade-off, while taking into account the selected configuration.

## 6.5. Scalable I/O and Visualization for Exascale Systems

### 6.5.1. CALCioM: mitigating cross-application I/O interference

**Participants:** Matthieu Dorier, Shadi Ibrahim, Gabriel Antoniu.

As larger supercomputers are used by an increasing number of applications in a concurrent manner, the interference produced by multiple applications accessing a shared parallel file system in contention becomes a major problem. Interference often breaks single-application I/O optimizations (such as access patterns preliminarily optimized to improve data locality on disks), thereby dramatically degrading application I/O performance, increasing run-time variability and, as a result, lowering machine-wide efficiency. We addressed this challenge by proposing CALCioM [15], a framework that aims to mitigate I/O interference through the dynamic selection of appropriate scheduling policies. CALCioM allows several applications running on a supercomputer to communicate and coordinate their I/O strategy in order to avoid interfering with one another. We examined four I/O strategies that can be accommodated in this framework: serializing, interrupting, interfering and coordinating. Experiments on Argonne's BG/P Surveyor machine and on several clusters of Grid'5000 showed that CALCioM can be used to improve the scheduling strategy efficiently and transparently between several otherwise interfering applications, given specified metrics of machine-wide efficiency. This work led to a publication at the IPDPS 2014 conference.

### 6.5.2. Omnisc'IO: Predicting the I/O patterns of HPC applications

**Participants:** Matthieu Dorier, Shadi Ibrahim, Gabriel Antoniu.

Many I/O optimizations including prefetching, caching, and scheduling, have been proposed to improve the performance of the I/O stack. In order to optimize these techniques, modeling and predicting spatial and temporal I/O patterns of HPC applications as they run, have become crucial. In this direction we introduced Omnisc'IO [16], an original approach that aims to make a step forward toward an intelligent I/O management of HPC applications in next-generation, post-Petascale supercomputers. It builds a grammar-based model of the I/O behavior of any HPC application, and uses this model to predict when future I/O operations will occur, as well as where and how much data will be accessed. Omnisc'IO is transparently integrated into the POSIX and MPI-I/O stacks and does not require any modification to application sources or to high-level I/O libraries. It works without prior knowledge of the application, and converges to accurate predictions within a couple of iterations only. Its implementation is efficient both in computation time and in memory footprint. Omnisc'IO was evaluated with four real HPC applications — CM1, Nek5000, GTC, and LAMMPS — using a variety of I/O backends ranging from simple POSIX to Parallel HDF5 on top of MPI-I/O. Our experiments showed that Omnisc'IO achieves from 79 % to 100 % accuracy in spatial prediction and an average precision of temporal predictions ranging from 0.2 seconds to less than a millisecond. This work was published at the SC14 conference and initiated the development of the Omnisc'IO software.

### *6.5.3. Smart In-Situ Visualization*

**Participants:** Lokman Rahmani, Matthieu Dorier, Gabriel Antoniu.

The increasing gap between computational power and I/O performance in new supercomputers has started to drive a shift from an offline approach to data analysis to an inline approach, termed *in-situ visualization* (ISV). While most visualization software now provides ISV, they typically visualize large dumps of unstructured data, by rendering everything at the highest possible resolution. This often negatively impacts the performance of simulations that support ISV, in particular when ISV is performed interactively, as in-situ visualization requires synchronization with the simulation. In this ongoing work, we investigate a smarter method of performing ISV. Our approach consists in adapting the resolution of regions of the visualization area based on how much their data are *relevant* with regards to the physical phenomena being simulated. In this direction, we first provide a generic definition of relevant data subsets based on *data variability*. Following this definition, we investigate various filtering algorithms to detect relevant data subsets automatically. The proposed filtering algorithms are derived from information theory, statistics and image processing. Our work is validated in the context of climate simulation, where we show an up to 40% improvement of time-to-solution without any significant loss regarding the quality of visualization (QoV). QoV loss is *quantified* using the structural similarity index metric (SSIM) that takes in consideration human visual system to compute visual errors.

## 6.6. Data Streaming and Small Data

### *6.6.1. JetStream: enabling high-performance event streaming across cloud data-centers*

**Participants:** Radu Tudoran, Alexandru Costan, Gabriel Antoniu.

The easily-accessible computation power offered by cloud infrastructures coupled with the revolution of Big Data are expanding the scale and speed at which data analysis is performed. In their quest for extracting value out of the 3 Vs of Big Data, applications process larger data sets, within and across clouds. Enabling fast data transfers across geographically distributed sites becomes particularly important for applications which manage continuous streams of events in real time. Scientific applications (e.g. the Ocean Observatory Initiative or the ATLAS experiment) as well as commercial ones (e.g. Microsoft's Bing and Office 365 large-scale services) operate on tens of data-centers around the globe and follow similar patterns: they aggregate monitoring data, assess the QoS or run global data mining queries based on inter-site event stream processing.

In [22] we propose a set of strategies for efficient transfers of events between cloud data-centers and we introduce JetStream: a prototype implementing these strategies as a high-performance, batch-based streaming middleware. JetStream is able to self-adapt to the streaming conditions by modeling and monitoring a set of context parameters. It further aggregates the available bandwidth by enabling multi-route streaming across cloud sites. The prototype was validated on tens of nodes from US and Europe data-centers of the Windows

Azure cloud using synthetic benchmarks and with application code in the context of the Alice experiment at CERN. The results show an increase in transfer rate of 250 times over individual event streaming. Besides, introducing an adaptive transfer strategy brings an additional 25 % gain. Finally, the transfer rate can further be tripled thanks to the use of multi-route streaming.

### *6.6.2. Efficient management of many small data objects*

**Participants:** Pierre Matri, Alexandru Costan, Gabriel Antoniu.

Large-scale intensive applications must often manage millions or even billions of small objects. Twitter, for example, has to record on average 5700 new tweets every second. Each of these objects are typically smaller than a kilobyte, and as a result, the database has to store billions of these objects. The sheer amount of objects and the small data sizes can also be found in many other applications, like sensor networks, or graph processing. Another important aspect are the access patterns of these applications where reads dominate over writes, which means the storage system has to be heavily optimized towards read performance.

To address these challenges, we are designing a novel storage system offering fast data access with minimal overhead. Learning from BlobSeer [33], we introduce a more efficient way to manage metadata. To this end, we propose to remove the centralised version manager and to distribute versions across the whole cluster using a distributed hash table. This greatly reduces the response times by allowing single-hop reads for most usage patterns. Additionally, this approach distributes the load over the whole cluster, thus providing a better horizontal scalability and fault tolerance.

**LAGADIC Project-Team**

# 6. New Results

## 6.1. Visual servoing

### *6.1.1. Photometric moment-based visual servoing*

**Participants:** Manikandan Bakthavatchalam, François Chaumette.

The goal of this work is to determine an adequate set of visual features to control the six degrees of freedom of a dynamic system. Thanks to a collaboration with Omar Tahri from Le2I in Le Creusot, we have been able to improve the results obtained previously with shifted moments for increasing the stability domain of visual servoing [24].

### *6.1.2. Histogram-based visual servoing*

**Participants:** Quentin Bateux, Eric Marchand.

Classically visual servoing considered the regulation in the image of a set of visual features (usually geometric features). Recently direct visual servoing schemes, such as photometric visual servoing, have been introduced in order to consider the image as a whole and thus avoid the extraction and the tracking of such geometric features. In this preliminary work, we propose a method to extend direct visual servoing approaches by using a global descriptor, namely intensity histograms, on the whole or multiple sub-sets of the images in order to achieve the control of a 6 degrees of freedom (DoF) robot.

### *6.1.3. Predictive visual servoing*

**Participants:** Nicolas Cazy, Paolo Robuffo Giordano, François Chaumette.

This study is devoted to the application of predictive control to visual servoing. In a first step, we have developed and compared several predictive models that can be useful when some visual features leave the camera field of view or are lost because of occlusions [25].

### *6.1.4. Visual servoing of cable-driven parallel robot*

**Participant:** François Chaumette.

This study is realized in collaboration with Rémy Ramadour and Jean-Pierre Merlet from Coprin group at Inria Sophia Antipolis. Its goal is to adapt visual servoing techniques for cable-driven parallel robot in order to achieve accurate manipulation tasks [46]. This study is in the scope of the Inria large-scale initiative action PAL (see Section 8.2.6 ).

### *6.1.5. Nanomanipulation*

**Participants:** Le Cui, Eric Marchand.

We began a work, within the ANR P2N Nanorobust project (see Section 8.2.1 ), on the development of micro- and nano-manipulation within SEM (Scanning Electron Microscope). Our goal is to provide visual servoing techniques for positioning and manipulation tasks with a nanometer precision. This year, we focused on the characterisation of the projection model of a SEM along with the approach required for its calibration [26]. We then address the problem of 6 dof control using photometric feature under an optical microscope [27]. Finally, we focused on the definition of control law able to control the motion along the Z axes with a SEM microscope. Indeed, considering that a SEM is subject to parallel projection model, motion along this axis is not observable. We then address this problem using defocus information. An autofocus process has also been studied.

### 6.1.6. Audio-based control

**Participants:** Aly Magassouba, François Chaumette.

This study is not concerned with visual servoing, but to the application of the same principle of closed loop control schemes to audio sensors. It is made in collaboration with Nancy Bertin from Panama group at Inria Rennes-Bretagne Atlantique. In a first step, we have determined the analytical form of the interaction matrix of audio features based on the time difference of arrival on two microphones. From this modeling step, we have determined the different virtual linkages that can be realized in function of the number and configuration of sources.

## 6.2. Visual navigation of mobile robots

### 6.2.1. Automous navigation of wheelchairs

**Participants:** Vishnu Karakkat Narayanan, François Pasteau, Marie Babel, François Chaumette.

The goal of this work is to design an autonomous navigation framework of a wheelchair by means of a single camera and visual servoing. We first focused on a corridor following task where no prior knowledge of the environment is required. The servoing process matches the non-holonomic constraints of the wheelchair and relies on two visual features, namely the vanishing point location and the orientation of the median line formed by the straight lines related to the bottom of the walls. This overcomes the initialization issue typically raised in the literature [9]. The control scheme has been implemented onto a robotized wheelchair and results show that it can follow a corridor with an accuracy of $\pm 3$ cm [16].

We then focused on a door passing task [33]. This doorway passing (and corridor turning) task employs the position of a single doorpost in the image as an input to a Lyapunov-based control scheme which allows the wheelchair to take up a desired trajectory about the doorpost. This trajectory avoids collision with the wall and guarantees that the wheelchair positions itself in front of the doorway regardless of its initial position. Results in simulation demonstrate the convergence and robustness of both control schemes. Experiments conducted on a wheelchair indicate the validity of applying the proposed low-level control system [16].

This study is in the scope of the Inria large-scale initiative action PAL (see Section 8.2.6 ) as well as of the Apash project (see Section 8.1.1 ).

### 6.2.2. Semi-autonomous control of a wheelchair for navigation assistance along corridors

**Participants:** Vishnu Karakkat Narayanan, Marie Babel, François Pasteau, Alexandre Krupa.

This study concerns a semi-autonomous control approach that we designed for safe wheelchair navigation. The control relies on the combination of primary tasks of wall avoidance as well as door passing performed by a dedicated visual servoing framework and a manual steering task. A smooth transition from manual driving to assisted navigation is obtained thanks to a gradual visual servoing activation method that guarantees the continuity of the control law. The visual servoing task is then progressively activated, when the wheelchair gets closer to the walls or doorposts, in order to avoid collisions [43]. Experimental results clearly show the ability of the approach to provide an efficient solution for wall avoiding and doorway passing purposes [58]. This study is in the scope of the Inria large-scale initiative action PAL (see Section 8.2.6 ) as well as of the Apash project (see Section 8.1.1 ). Tests with disabled patients in the rehabilitation center Pôle Saint Hélier (Rennes) are under progress and first results prove the ability of our system to smoothly correct the trajectory of the wheelchair in case of hazardous situations.

Current research works are oriented towards multimodal sensor-based servoing, as well as haptic feedback that leads to an intuitive assistive wheelchair navigation. This work is realized in collaboration with Maud Marchal (Hybrid team). In addition, we are currently working with e-Motion team to design a vision-based human-aware semi-autonomous navigation system.

### 6.2.3. Social Spacing and human-robot interaction

**Participants:** Panagiotis Papadakis, Patrick Rives.

A novel probabilistic framework was introduced capable of instantiating diverse models of social spacing and accounting for distinctive dimensions in human-robot interaction, namely, perception capacity and certainty [42]. We have concretely shown how our method allows smooth adaptation in the situation awareness of a robot within common human-robot interaction examples and further showed its utility at the level of path planning by adapting trajectories to social sensitivity levels.

This approach is currently extended to take into account human activity cues in order to enhance robot mapping and navigation and in particular in filtering noisy human detections, detecting passages such as doors and staircases, inferring space occupancy and allowing navigation within unexplored areas.

### *6.2.4. Target tracking*

**Participants:** Ivan Markovic, François Chaumette.

This study was realized in the scope of the FP7 Regpot Across project (see Section 8.3.1.2 ) during the three-month visit of Ivan Markovic, Ph.D. student at the University of Zagreb. It consisted in developing a pedestrian visual tracking from an omni-directional fish-eye camera and a visual servoing control scheme so that a mobile robot is able to follow the pedestrian [37]. This study has been validated on our Pioneer robot (see Section 5.4.2 ).

### *6.2.5. Obstacle avoidance*

**Participants:** Suman Raj Bista, Fabien Spindler, François Chaumette.

This study was realized in collaboration with Andrea Cherubini who is now Assistant Prof. at Université de Montpellier. It is concerned with our long term researches about visual navigation from a visual memory without any accurate 3D localization [4]. In order to deal with obstacle avoidance while preserving the visibility in the visual memory, we have proposed a control scheme based on tentacles for fusing the data provided by a pan-tilt camera and a laser range sensor [11]. A new study devoted to indoors navigation from segments has started recently.

## 6.3. Visual tracking and state estimation

### *6.3.1. 3D model-based tracking*

**Participant:** Eric Marchand.

This study focused on the issue of estimating the complete 3D pose of the camera with respect to a potentially textureless object, through model-based tracking. We proposed to robustly combine complementary geometrical and color edge-based features in the minimization process, and to integrate a multiple-hypotheses framework in the geometrical edge-based registration phase [45]. This method will be tested in the scope of the FP7 RemoveDebris project [36].

### *6.3.2. Pose estimation through plane tracking*

**Participants:** Aurélien Yol, Eric Marchand.

We proposed a method for localizing an Unmanned Aerial Vehicle (UAV) using georeferenced aerial images. Here we provide a multiple usage localization algorithm based on vision only. To ensure robustness, we choose to use the Mutual Information (MI) within a dense tracking process. MI proved to be very robust toward local and global scene variations. However, dense approaches are often related to drift disadvantages. We solve this problem by using georeferenced images. The localization algorithm has been demonstrated through the localization of a hexarotor UAV fitted with a downward looking camera during real flight tests [53].

### *6.3.3. 3D tracking of deformable objects*

**Participants:** Bertrand Delabarre, Eric Marchand.

We consider the problem of dense non-rigid visual tracking robust towards global illumination perturbations of the observed scene. The similarity function is based on the sum of conditional variance (SCV). With respect to most approaches that minimize the sum of squared differences, which is poorly robust towards illumination variations in the scene, the choice of SCV as our registration function allows the approach to be naturally robust towards global perturbations. Moreover, a thin-plate spline warping function is considered in order to take into account deformations of the observed template [28].

### 6.3.4. Structure from motion

**Participants:** Riccardo Spica, Paolo Robuffo Giordano, François Chaumette.

Structure from motion (SfM) is a classical and well-studied problem in computer and robot vision, and many solutions have been proposed to treat it as a recursive filtering/estimation task. However, the issue of *actively* optimizing the transient response of the SfM estimation error has not received a comparable attention. In the work [18], we showed how to design an online active SfM scheme characterized by an error transient response equivalent to that of a reference linear second-order system with desired poles. Indeed, in a nonlinear context, the observability properties of the states under consideration are not (in general) time-invariant but may depend on the current state and on the current inputs applied to the system. It is then possible to simultaneously act on the estimation gains and system inputs (i.e., the camera velocity for SfM) in order to optimize the observation process and impose a desired transient response to the estimation error. The theory has a general validity and can be applied to many different contexts such as when dealing with point features [18], solid objects like spheres or cylinders [51], or planar regions [47]. Furthermore, the active SfM scheme can also be embedded within a classical visual servoing law exploiting the redundancy of the camera motion w.r.t. the considered visual task [52].

### 6.3.5. Robust visual odometry

**Participants:** Tawsif Gokhool, Patrick Rives, Renato José Martins.

Our aim is concentrated around building ego-centric topometric maps represented as a graph of salient keyframe nodes [14]. Additionally, visual odometry from frame to keyframe alignment helps significantly in drift reduction. On the other hand, the sparsity in this kind of graphical representation leads to reduced overlapping between keyframes which can degrade localisation robustness. Our chosen spherical $360^0$ field of view (FOV) configuration alleviates the overlapping issue by providing an enriched model of the environment with photometric and geometric information content. Following a multitude of advantages with information fusion, merging of frames in a single representation deals with the problem of data redundancy and sensor noise suppression.

Therefore, the second fold of this work consisted in improving the identified conceptual loopholes above by first proposing a generic uncertainty propagation model as applied to our spherical RGB-D database. Secondly, a probabilistic framework was derived which led to a Mahalanobis inconsistency test incorporating both geometric and photometric uncertainty models [32]. Our framework was further improved by adding up a probabilistic model to filter out dynamic points temporally. Finally, the entire probabilistic framework was applied in order to track the most stable points over time.

## 6.4. 3D Scene Mapping

### 6.4.1. New RGB-D sensor design for indoor 3D mapping

**Participants:** Eduardo Fernandez Moral, Patrick Rives.

A multi-sensor device has been developed for omnidirectional RGB-D (color+depth) image acquisition (see Fig. 3 .a). This device allows to acquire such omnidirectional images at high frame rates (30 Hz). This approach has advantages over other alternatives used today in terms of accuracy and real-time spherical image construction for indoor environments, which are specially interesting for mobile robotics. This device has important prospective applications as fast 3D-reconstruction or Slam.

A calibration method for such device was developed [31], which takes into account the bias of each sensor independently. The proposed calibration method does not require any specific calibration pattern, taking into account the planar structure from the scene to cope with the fact that there is no overlapping between sensors.

In a first instance, this sensor has been exploited for localization and mapping research with mobile robots. For that, the sensor is mounted on a mobile platform together with an standard computer (see Fig. 3 .a). A method to perform image registration and visual odometry has been developed. This method relies in the matching of planar primitives that can be efficiently obtained from the depth images. This technique performs considerably faster than previous registration approaches like ICP, or dense photoconsistency alignment. These last achieve however a better accuracy than our method, what suggests that our method can be used as an initial step to speed-up those.

Slam is also addressed with this device. A solution to this problem using our omnidirectional RGB-D sensor is being researched. The ongoing experiments have shown some initial results for metric-topological pose-graph Slam, where the map consists of a set of spherical keyframes, which are located in a topological arrangement according to their shared observations.

### 6.4.2. Compact 3D scene representation

**Participants:** Renato José Martins, Patrick Rives, Tawsif Gokhool.

This work follows in the direction of precise and compact scene representation of large scale environments. The aim is to build a complete geometric and photometric "minimal" model, which is stored within a sparse set of augmented spherical images to asset photo-geometry consistence of the scene from multiple points-of-views. In this direction, an uncertainty model from the full structure combined with those of poses was proposed for point-to-point egocentric fusion. This model allows to reduce sensor noise in a particular keyframe sphere when performing a multi-frame fusion scheme of coherent near information. This first fusion scheme is then improved by exploiting the rigidity/influence of neighboring points representing the surface. For that, an intermediary higher level abstraction of the point cloud is generated by partitioning the input domain into elementary cells, then reducing the number of degrees of freedom and enforcing constraints over the points segmented as being part of the same surface.

The adopted solution is a "weaker" representation of a 3D boundary mesh, based on discontinuous convex planar patches, with the segmentation being done considering the geometry (region growing) or photometry (SLIC superpixels). This synthetic scene built with the planar geometric police proved to well represent the original scene (for both indoor and outdoor real data) with a significant small amount of patches and it is exploited to build robust useful "dynamic" 4D world model, which in turn can be used for assisted/autonomous navigation or virtual reality applications.

### 6.4.3. Semantic mapping

**Participants:** Romain Drouilly, Patrick Rives, Panagiotis Papadakis.

Autonomous navigation is one of the most challenging problems to address to allow robots to evolve in our everyday environments. Map-based navigation has been studied for a long time and researches have produced a great variety of approaches to model the world. However, semantic information has only recently been taken into account in those models to improve robot efficiency [56]. The goal of this work is to study how semantics can be used to improve all the steps of navigation process. In a first time, we have developed a new navigation-oriented hybrid metric-topological-semantic model of the world. It captures high-level information and uses it to build extremely compact description of large environments. Then we have used it to design an efficient localization algorithm, able to find a given map content faster than classical methods and allowing human-understandable queries [30]. In a second time, we have studied how semantics can be used to discover unobserved things in the scene. Particularly, we have shown that both statics and dynamic entities, identified by a robot, can inform about the structure of the environment in unobserved areas [29] . We have used this to do "map extrapolation", that is extending a map beyond robot's perceptual limits by reasoning on semantics. This approach has been shown to be of great interest in everyday-life environment. Finally, we have proposed a new scheme for trajectory planing, taking into account not only geometric constraints

but also high-level understanding of the world. We have shown the usefulness of this approach to navigate complex environments with highly dynamic areas on both simulated and real-world datasets, well-suited for large outdoor environment navigation.

### 6.4.4. Augmented reality

**Participant:** Eric Marchand.

Using Slam methods becomes more and more common in Augmented Reality (AR). To achieve real-time requirement and to cope with scale factor and the lack of absolute positioning issue, we proposed to decouple the localization and the mapping step. This approach has been validated on an Android Smartphone through a collaboration with Orange Labs [38][39]

## 6.5. Medical robotics

### 6.5.1. Non-rigid target tracking in ultrasound images based on hierarchical grid interpolation

**Participants:** Lucas Royer, Jason Chevrie, Marie Babel, Alexandre Krupa.

In order to track the motion of a tumour or cyst during needle insertion, we developed a first approach to track a deformable target within a sequence of 2D ultrasound images. It is based on a dedicated hierarchical grid interpolation algorithm (HGI) that is typically used for real-time video compression purposes. This approach provides a continuous motion representation of the target by using a grid of control points that models both their global displacement and local deformations. The motion of each control point is estimated by a hierarchical and multi-resolution local search method in order to minimize the sum of squared difference of the target pixel intensity between successive images. This approach was validated from 2D ultrasound images of real human tissues undergoing rigid and non-rigid deformations [48] and was recently adapted for tracking 3D deformations.

### 6.5.2. Non-rigid target tracking in ultrasound images based on physically-based model

**Participants:** Lucas Royer, Alexandre Krupa.

A second approach for automatically tracking deformable target within 2D ultrasound images has been developed [50]. It combines dense information with a physically-based model and has therefore the advantage of not using any fiducial marker. The physical model is represented by a mass-spring damper system driven by external and internal forces. The external forces are obtained by maximizing an image similarity metric between a reference target and the deformed target along the time. The internal forces of the mass-spring damper system constrain the deformation to be physically plausible and therefore efficiently reduce the sensitivity to the speckle noise. This second approach was validated on simulated and real data, both for rigid and non-rigid motions of soft tissues [49]. It was recently extended for deformable target tracking in 3D ultrasound volumes.

### 6.5.3. 3D steering of flexible needle by visual servoing

**Participants:** Alexandre Krupa, Pierre Chatelain.

The objective of this work is to provide robotic assistance during needle insertion procedures such as biopsy or ablation of localized tumor. A method has been developed for steering a beveled-tip flexible needle actuated by a robotic arm in such a way to control the needle curvature in 3D space [34]. It is based on the design of a new duty-cycling control strategy that makes possible to control both the 2 lateral angular velocities and the insertion velocity of the needle tip (3 DOF). An image-based visual servoing approach has then been developed to automatically position the needle tip on a 3D target indicated by the user. It is based on the use of geometrical visual features extracted from 2 images provided by 2 orthogonal cameras observing a translucent gelatin phantom where the needle is inserted. Preliminary results of this automatic targeting task demonstrate the feasibility of this new concept and its robustness to needle kinematic model errors [35]. We recently extended this approach to automatically steer the needle toward a target by an image-based visual servoing that uses geometrical features extracted from images provided by a 3D ultrasound probe.

### 6.5.4. Optimization of ultrasound image quality by visual servoing

**Participants:** Pierre Chatelain, Alexandre Krupa.

This study focuses on the automatic positioning of a 2D ultrasound probe in such a way to optimize the quality of the acquired ultrasound images. It is based on the recent framework of ultrasound confidence map, developed in the Chair for Computer Aided Medical Procedures and Augmented Reality of Prof. Nassir Navab, which aims at estimating the per-pixel quality of the ultrasound signal based on a model of sound propagation in soft tissues. In collaboration with Nassir Navab we considered this ultrasound confidence map as a new modality and recently designed a visual servoing control law for image quality optimization.

## 6.6. Control of single and multiple Unmanned Aerial Vehicles

### 6.6.1. State estimation and flight control of quadrotor UAVs

**Participant:** Paolo Robuffo Giordano.

Over the last years the robotics community witnessed an increasing interest in the Unmanned Aerial Vehicle (UAV) field. In particular quadrotor UAVs have become more and more widespread in the community as experimental platform for, e.g., testing novel 3D planning, control and estimation schemes in real-world indoor and outdoor conditions. Indeed, in addition to being able to take-off and land vertically, quadrotors can reach high angular accelerations thanks to the relatively long lever arm between opposing motors. This makes them more agile than most standard helicopters or similar rotorcraft UAVs, and thus very suitable to realize complex tasks such as aerial mapping, air pollution monitoring, traffic management, inspection of damaged buildings and dangerous sites, as well as agricultural applications such as pesticide spraying.

Despite these clear advantages, a clear shortcoming of the quadrotor design lies in its inherent underactuation (only 4 actuated propellers for the 6 dofs of the quadrotor pose). This underactuation limits the quadrotor flying ability in free or cluttered space and, furthermore, it also degrades the possibility of interacting with the environment by exerting desired forces in arbitrary directions. In [17], a novel design for a quadrotor UAV with tilting propellers which is able to overcome these limitations has been presented and experimentally validated. Indeed, the additional set of 4 control inputs actuating the propeller tilting angles can be shown to yield full actuation to the quadrotor position/orientation in space, thus allowing it to behave as a fully-actuated flying vehicle and to overcome the aforementioned underactuation problem.

This work has been realized in collaboration with the Max Planck Institute for Biological Cybernetics, Tübingen, Germany.

### 6.6.2. Collective control of multiple UAVs

**Participant:** Paolo Robuffo Giordano.

The challenge of coordinating the actions of multiple robots is inspired by the idea that proper coordination of many simple robots can lead to the fulfilment of arbitrarily complex tasks in a robust (to single robot failures) and highly flexible way. Teams of multi-robots can take advantage of their number to perform, for example, complex manipulation and assembly tasks, or to obtain rich spatial awareness by suitably distributing themselves in the environment. Within the scope of robotics, autonomous search and rescue, firefighting, exploration and intervention in dangerous or inaccessible areas are the most promising applications.

In the context of multi-robot (and multi-UAV) coordinated control, *connectivity* of the underlying graph is perhaps the most fundamental requirement in order to allow a group of robots accomplishing common goals by means of *decentralized* solutions. In fact, graph connectivity ensures the needed continuity in the data flow among all the robots in the group which, over time, makes it possible to share and distribute the needed information. However, connectivity alone is not sufficient to perform certain tasks when only *relative sensing* is used. For these systems, the concept of *rigidity* provides the correct framework for defining an appropriate sensing and communication topology architecture. Rigidity is a combinatorial theory for characterizing the "stiffness" or "flexibility" of structures formed by rigid bodies connected by flexible linkages or hinges. In a broader context, rigidity turns out to be an important architectural property of many multi-agent systems when

a common inertial reference frame is unavailable. Applications that rely on sensor fusion for localization, exploration, mapping and cooperative tracking of a target, all can benefit from notions in rigidity theory. The concept of rigidity, therefore, provides the theoretical foundation for approaching decentralized solutions to the aforementioned problems using distance measurement sensors, and thus establishing an appropriate framework for relating system level architectural requirements to the sensing and communication capabilities of the system.

In [22], a decentralized gradient-based rigidity maintenance action for a group of quadrotor UAVs has been proposed and tested in real experimental conditions. By starting in a rigid configuration, the group of UAVs is able to estimate their relative position from sole relative distance measurements, and then use these estimated relative positions in a control action able to preserve rigidity of the whole formation despite presence of sensor limitations (maximum range and line-of-sight occlusions), possible collisions with obstacles and inter-robot collisions. Furthermore, in [54] the rigidity theory has been extended to the case of *bearing measurements*, and directed graphs.

These works were realized in collaboration with the robotics group at the Max Planck Institute for Biological Cybernetics, Tübingen, Germany and with Technion, Israel.

**LINKMEDIA Project-Team**

# 6. New Results

## 6.1. Highlights of the Year

BEST PAPER AWARD :

[21] **European Symposium on Research in Computer Security**. R. BALU, T. FURON, S. GAMBS.

## 6.2. Unsupervised motif discovery

### *6.2.1. Clustering by diverting supervised machine learning*

**Participants:** Vincent Claveau, Patrick Gros, Abir Ncibi.

Knowledge discovery aims at bringing out coherent groups of objects and is usually based on clustering which necessitates defining a notion of similarity between objects. In practice, this strong prior is often neither possible nor welcome. We proposed to divert supervised machine learning (ML) techniques in order to calculate, indirectly and without supervision, similarities among objects. Our approach consists in generating artificial labeling problems on the data to reveal regularities between objects through their labeling. In [28], we show how this framework can be implemented and experimented on two information extraction/discovery tasks concerned with named entities. The ML technique diverted to exhibit similarities between with the named entities are the Conditional Random Fields. This same method can also be applied with less common ML techniques: In [59], we show that Inductive Logic Programming can also be used to cluster complex data. Thanks to the ability of ILP to handle data that cannot be expressed under the usual attribute-value representation, we use it to make emerge clusters of TV broadcasts based only on their broadcasting information (date, time, length, etc.).

### *6.2.2. Spoken term discovery applied to audio thumbnailing*

**Participants:** Sébastien Campion, Guillaume Gravier.

We evaluated a system to create audio thumbnails of spoken content, i.e., short audio summaries representative of the entire content, without resorting to a lexical representation. As an alternative to searching for relevant words and phrases in a transcript, unsupervised motif discovery is used to find short, word-like, repeating fragments at the signal level without acoustic models. The output of the word discovery algorithm is exploited via a maximum motif coverage criterion to generate a thumbnail in an extractive manner. A limited number of relevant segments are chosen within the data so as to include the maximum number of motifs while remaining short enough and intelligible. Evaluation is performed on broadcast news reports with a panel of human listeners judging the quality of the thumbnails. Results indicate that motif-based thumbnails stand between random thumbnails and ASR-based keywords, however still far behind thumbnails and keywords humanly authored [34].

### *6.2.3. Unsupervised video structure mining with grammatical inference*

**Participants:** Guillaume Gravier, Bingqing Qu.

*In collaboration with Jean Carrive and Félicien Vallet, Institut National de l'Audiovisuel.*

Unsupervised approaches were introduced a few years ago to analyze the structure of TV programs, relying on the discovery of repeated elements within a program or across multiple episodes of the same program. These methods can discover key repeating elements, such as jingles and separators, however they cannot infer the entire structure of a program. In [48], we studied a hierarchical use of grammatical inference to yield a temporal grammar of a program from a collection of episodes, discovering both the vocabulary of the grammar and the temporal organization of the words from the vocabulary. Using a set of basic event detectors and simple filtering techniques to detect repeating elements of interest, a symbolic representation of each episode is derived based on minimal domain knowledge. Grammatical inference based on multiple sequence alignment is then used in a hierarchical manner to provide a temporal grammar of the program at various levels of details.

### 6.2.4. Efficient indexing for content retrieval

**Participants:** Raghavendran Balu, Teddy Furon, Hervé Jégou.

*In collaboration with Miajing Shi during, visiting Ph. D. student from Pekin University.*

Many nearest neighbor search algorithms rely on encoding real vectors into binary vectors. The most common strategy projects the vectors onto random directions and takes the sign to produce so-called sketches. In [22], we discuss the sub-optimality of this choice, and propose a better encoding strategy based on the quantization and reconstruction points of view. Our second contribution is a novel asymmetric estimator for the cosine similarity. Similar to previous asymmetric schemes, the query is not quantized and the similarity is computed in the compressed domain. We tackled the same similarity estimation problem with a rather different approach in [52], where we assume that only a few vectors of the database, so-called heavy hitters, have a similarity to the query that significantly deviates from 0. For this purpose, we have introduced a group testing framework for detecting large similarities between high-dimensional vectors, such as descriptors used in state-of-the-art description of multimedia documents. We produce a set of group representations that jointly encode several vectors into a single one, in the spirit of group testing approaches. By comparing a query vector to several of these intermediate representations, we screen the large values taken by the similarities between the query and all the vectors, at a fraction of the cost of exhaustive similarity calculation. Unlike concurrent indexing methods that suffer from the curse of dimensionality, our method exploits the properties of high-dimensional spaces.

## 6.3. Multimedia content description and structuring

### 6.3.1. Linguistic knowledge extraction

#### 6.3.1.1. Identifying events in texts

**Participant:** Vincent Claveau.

*In collaboration with Béatrice Arnulphy, former team member now with ANR, Xavier Tannier and Anne Vilnat, LIMSI.*

Identifying events from texts is an information extraction task necessary for many NLP applications. Through the TimeML specifications and TempEval challenges, it has received some attention in the last years, yet, no reference result was available for French. In [20], we tried to fill this gap by proposing several event extraction systems, combining for instance Conditional Random Fields, language modeling and k-nearest-neighbors. These systems are evaluated on French corpora and compared with state-of-the-art methods on English. The very good results obtained on both languages validate our whole approach and set new standard for French.

#### 6.3.1.2. Morpho-semantic analysis of terms

**Participants:** Vincent Claveau, Ewa Kijak.

In most Indo-European languages, many biomedical terms are rich morphological structures composed of several constituents mainly originating from Greek or Latin. The interpretation of these compounds are keystones to access information. Following our work on morphology in the biomedical domain, we proposed different techniques to generate probabilistic morph-semantic resources and we show how these alignment information can be used for segmenting compounds, attaching semantic interpretation to each part, proposing definitions (gloses) of the compounds [26]. When possible, these tasks are compared with state-of-the-art tools, and the results show the interest of our automatically built probabilistic resources.

#### 6.3.1.3. Distributional semantics

**Participants:** Vincent Claveau, Ewa Kijak.

*In collaboration with Olivier Ferret, CEA-LIST.*

We addressed the issue of building and improving a distributional thesaurus. We first show that existing tools from the information retrieval domain can be directly used in order to build a thesaurus with state-of-the-art performance. Secondly, we focus more specifically on improving the obtained thesaurus, seen as a graph of k-nearest neighbors. By exploiting information about the neighborhood contained in this graph, we propose several contributions. 1) We show how the lists of neighbors can be globally improved by examining the reciprocity of the neighboring relation, that is, the fact that a word can be close to another and vice-versa. 2) We also propose a method to associate a confidence score to any lists of nearest neighbors (i.e., any entry of the thesaurus). 3) Last, we demonstrate how these confidence scores can be used to reorder the closest neighbors of a word. These different contributions are validated through experiments and offer significant improvement over the state-of-the-art [27], [60].

### 6.3.2. OCR and speech content-based description

#### *6.3.2.1. Use of stress information for robust speech recognition*

**Participant:** Guillaume Gravier.

*In collaboration with S. Ziegler, PANAMA Inria team and Laboratoire de Sciences Cognitives et Psycholinguistique.*

[44] presents a study on the robustness of stress information for automatic speech recognition in the presence of noise. The syllable stress, extracted from the speech signal, was integrated in the recognition process by means of a previously proposed decoding method. Experiments were conducted for several signal-to-noise ratio conditions and the results show that stress information is robust in the presence of medium to low noise. This was found to be true both when syllable boundary information was used for stress detection and when this information was not available. Furthermore, the obtained relative improvement increased with a decrease in signal quality, indicating that the stressed parts of the signal can be considered islands of reliability.

#### *6.3.2.2. Boosting bonsai trees for handwritten/printed text discrimination*

**Participant:** Christian Raymond.

*In collaboration with Yann Ricquebourg, Baptiste Poirriez, Aurélie Lemaitre and Bertrand Coüasnon, IRISA.*

Boosting over decision-stumps proved its efficiency in natural language processing, essentially with symbolic features, and its good properties (fast, few and not critical parameters, not sensitive to overfitting) could be of great interest in the numeric world of pixel images. In [51], we investigated the use of boosting over small decision trees in image classification processing for the discrimination of handwritten/printed text. We conducted experiments to compare with usual SVM-based classification revealing convincing results with very close performance, but with faster predictions and behaving far less as a black-box. Those promising results tend to make use of this classifier in more complex recognition tasks like multiclass problems.

#### *6.3.2.3. Speaker role detection from spoken document*

**Participant:** Christian Raymond.

*In collaboration with LIMSI and LIUM.*

In [40] and [41], we tackle the problem of speaker role detection in broadcast news shows. In the literature, many proposed solutions are based on the combination of various features coming from acoustic, lexical and semantic information with a machine learning algorithm. Many previous studies mention the use of boosting over decision stumps to combine efficiently these features. We proposed a modification of this state-of-the-art machine learning algorithm changing the weak learner (decision stumps) by small decision trees, denoted bonsai trees. Experiments show that using bonsai trees as weak learners for the boosting algorithm largely reduces both system error rate and learning time.

### 6.3.3. Image and video description and classification

#### *6.3.3.1. Fine-grain image classification*

**Participants:** Teddy Furon, Philippe-Henri Gosselin, Hervé Jégou.

*In collaboration with Xerox Research Center Europe.*

We have addressed the problem of instance classification: our goal is to annotate images with tags corresponding to objects classes which exhibit small intra-class variations such as logos, products or landmarks. Our first contribution on image classification [13] describes the processing pipeline, which has won FGCOMP challenge associated with Imagenet. It improves a standard method based on Fisher vectors to adapt it to the context of fine-grained classes, where the difference between classes rely on few but typical visual differences. On the same task, we have proposed a novel algorithm [39] for the selection of class-specific *prototypes* which are used in a voting-based classification scheme.

*6.3.3.2. Aggregation of local descriptors*

**Participants:** Teddy Furon, Hervé Jégou, Giorgos Tolias.

*In collaboration with the University of Oxford.*

For unsupervised particular object and image recognition, we have considered the design of a single vector representation for an image that embeds and aggregates a set of local patch descriptors such as SIFT. In [36], we make two contributions, both aimed at regularizing the individual contributions of the local descriptors in the final representation. The first is a novel embedding method that avoids the dependency on absolute distances by encoding directions. The second contribution is a "democratization" strategy that further limits the interaction of unrelated descriptors in the aggregation stage. In [36], we addressed another issue inherent to existing encoding algorithms: Image search systems based on local descriptors typically achieve orientation invariance by aligning the patches on their dominant orientations. This choice introduces too much invariance because it does not guarantee that the patches are rotated consistently. To address this problem, we have introduced another aggregation strategy of local descriptors that achieves this covariance property by jointly encoding the angle in the aggregation stage in a continuous manner. It is combined with an efficient monomial embedding to provide a codebook-free method to aggregate local descriptors into a single vector representation.

*6.3.3.3. Action localization in videos*

**Participants:** Mihir Jain, Hervé Jégou.

*In collaboration with the University of Amsterdam and the project-team SERPICO.*

We have tackled the problem of action localization in videos [35], where the objective is to determine when and where certain actions appear. We introduce a sampling strategy, called tubelets and inspired a method recently introduced for image detection. It drastically reduces the number of hypotheses that are likely to include the action of interest. By using super-voxels and employing a criterion that reflects how action related motion deviates from background motion, the method is specifically adapted to 2D+t sequences and establishes the new state-of-the-art for action localization on the public datasets UCF Sports and MSR-II.

### *6.3.4. Text description for information retrieval*

**Participants:** Vincent Claveau, Sébastien Le Maguer.

*In collaboration with Natalia Grabar, STL UMR8163, and Thierry Hamon, LIMSI*

Following previous work, we investigated the interest of "bag of bags of features" representation for texts in an vector-space information retrieval setting. Each text is thus represented as a bag of vector. With this representation, computing the similarity between two texts necessitates to aggregate every vector to vector similarity for the two bags. In [58], we examine the expected properties of such an aggregation function and show their influence through different experiments. When some specific conditions are met, we show that the gains over standard representation can be very important.

With a team composed with members of TEXMEX/LINKMEDIA, LIMSI and STL, we have participated to the biomedical information retrieval challenge proposed in the framework of CLEF eHealth [25]. For this first participation, our approach relies on a state-of-the-art IR system called Indri, based on statistical language modeling, and on semantic resources. The purpose of semantic resources and methods is to manage the term variation such as synonyms, morpho-syntactic variants, abbreviation or nested terms. Different combinations of resources and Indri settings are explored, mostly based on query expansion. We obtained good overall results (3rd in terms of MAP) and confirmed the interest of query expansion to retrieve a maximum of relevant documents.

## 6.4. Linking, navigation and analytics

### 6.4.1. NLP-driven hyperlink construction in broadcast videos

**Participants:** Rémi Bois, Vincent Claveau, Guillaume Gravier, Pascale Sébillot, Anca-Roxana Şimon.

*In collaboration with Sien Moens, Katholieke Universiteit Leuven, Éric Jamet and Martin Ragot, Univ. Rennes 2.*

The hyperlinking sub-task of the MediaEval Search and Hyperlinking task aims at creating hyperlinks between predefined anchor segments, i.e., fragments of videos, and short related video segments, called targets, that have to be automatically extracted from videos of a large collection. Capitalizing on the experience acquired in previous participations [54], we proposed a two step approach exploiting speech material: Potential target segments are first generated relying on a topic segmentation technique; For each anchor, the best 20 target segments are then selected according to two distinct strategies. The first strategy focuses on the identification of very similar targets using n-grams and named entities, while the second one makes use of an intermediate structure built from topic models, which offers the possibility to control serendipity and to explain the links created [53].

In 2014, we also initiated the CominLabs project "Linking media in acceptable hypergraphs" dedicated to the creation of explicit and meaningful links between multimedia documents or fragments of documents. Two main issues were adressed: The construction of a corpus, composed of audio and video news, reports and debates, newspapers and blog websites, as well as social networks; A preliminary study of the perceived usefulness of various types of links by end-users.

### 6.4.2. Analytics in collections of art critics

**Participant:** Vincent Claveau.

*In collaboration with Fabienne Moreau and Nicolas Thély, Univ. Rennes 2.*

We aim at exploiting text mining techniques in the service of digital humanities, and more precisely in the field of art criticism. It relies on a collaboration between our team, linguists and art and aesthetics specialists. In preliminary work [56], we adapted term extraction, named entity recognition and information retrieval techniques to this field to extract multiple linguistic clues from art review articles. Future work will make the most of these clues and clustering approaches to build a navigable and structured collection of the articles.

### 6.4.3. Data models for navigation

**Participant:** Laurent Amsaleg.

*In collaboration with Björn Þór Jónsson, Grímur Tómasson, Hlynur Sigurþórsson, Áslaug Eríksdóttir and Marta Kristin Larusdottir, School of Computer Science, Reykjavík University.*

Digital photo collections—personal, professional, or social—have been growing ever larger, leaving users overwhelmed. It is therefore increasingly important to provide effective browsing tools for photo collections. Learning from the resounding success of multi-dimensional analysis (MDA) in the business intelligence community for On-Line Analytical Processing (OLAP) applications, we propose a multi-dimensional model for media browsing, called $M^3$, that combines MDA concepts with concepts from faceted browsing. We present the data model and describe preliminary evaluations, made using server and client prototypes, which indicate that users find the model useful and easy to use [38]. A photo navigation prototype was demonstrated at the Intl. Conf. on Multimedia Modeling [37].

### 6.4.4. Exploiting k-nn graphs for image retrieval

**Participants:** Laurent Amsaleg, Hervé Jégou, Giorgos Tolias.

We have proposed two techniques exploiting the relationship between the images with a collection. In [29], we revisit how to exploit the $k$-reciprocal nearest neighbors to produce, for a given query, a neighborhood that improves over the one obtained with the original metric. This strategy is simpler than concurrent prior work, yet it is both effective and less sensitive to parameters. Second, we propose to employ measures defined on sets of shared nearest neighbors in order to re-rank the shortlist. Both methods are simple, yet they significantly improve the accuracy of image search engines on standard benchmarks. We also introduced a query expansion technique [18] for image search that is faster and more precise than the existing ones. The expansion generates an enriched representation which refines the initial local descriptors *individually* by aggregating those of the database, while new descriptors are produced from the images that are deemed relevant. The technique has two computational advantages over other query expansion techniques. First, the size of the enriched representation is comparable to that of the initial query. Second, the technique is effective even without using any geometry, in which case searching a database comprising 105k images typically takes 80 ms on a desktop machine. Overall, our technique significantly outperforms the visual query expansion state of the art on popular benchmarks.

## 6.5. Content-based information retrieval and security

### 6.5.1. Analysis of privacy preserving data aggregation for recommendation systems

**Participants:** Raghavendran Balu, Teddy Furon.

*Work under the Alcaltel-Lucent / Inria common Lab, in collaboration with Armen Aghasaryan, Dimitre Davidov, Makram Bouzid (ALU) and Sébastien Gambs (Inria Rennes Cidre team project).*

We consider personalized recommendation systems in which before publication, the profile of a user is sanitized by a non-interactive mechanism compliant with the concept of differential privacy. We analyze two existing schemes offering a differentially private representation of profiles: BLIP (BLoom-and-flIP) and JLT (Johnson-Lindenstrauss Transform). For assessing their security levels, we play the role of an adversary aiming at reconstructing a user profile [21]. We compare two inference attacks, namely single and joint decoding. The first one decides of the presence of a single item in the profile, and sequentially explores all the item set. The second one decides whether a subset of items is likely to be the user profile, and considers all the possible subsets. Our contributions are a theoretical analysis as well as a practical implementation of both attacks, which were evaluated on datasets of real user profiles. The results obtained clearly demonstrates that joint decoding is the most powerful attack, while also giving useful insights on how to set the differential privacy parameter $\epsilon$.

### 6.5.2. Content based image retrieval with privacy

**Participants:** Laurent Amsaleg, Teddy Furon, Li Weng.

*Work initiated during a collaboration with A. Morton, L. Weng (with* LINKMEDIA *since May 2014) and S. Marchand-Maillet, Université de Genève.*

We propose a privacy protection framework for large-scale content-based information retrieval. It offers two layers of protection. First, robust hash values are used as queries to prevent revealing original content or features. Second, the client can choose to omit certain bits in a hash value to further increase the ambiguity for the server. Due to the reduced information, it is computationally difficult for the server to know the client's interest. The server has to return the hash values of all possible candidates to the client. The client performs a search within the candidate list to find the best match. Since only hash values are exchanged between the client and the server, the privacy of both parties is protected. We introduce the concept of tunable privacy, where the privacy protection level can be adjusted according to a policy. It is realized through hash-based piece-wise inverted indexing. The idea is to divide a feature vector into pieces and index each piece with a sub-hash value. Each sub-hash value is associated with an inverted index list. The framework has been extensively tested using a large image database. We have evaluated both retrieval performance and privacy-preserving performance for a particular content identification application.

### 6.5.3. Privacy protection in face recognition

**Participant:** Teddy Furon.

*In collaboration with B. Bhattarai, A. Mignon, F. Jurie, GREYC, Université of Caen.*

We investigated a new approach for de-identifying face images, *i.e.* for preventing automatic matching with public face collections. The overall motivation is to offer tools for privacy protection on social networks. We address this question by drawing a parallel between face de-identification and oracle attacks in digital watermarking. In our case, the identity of the face is seen as the watermark to be removed. Inspired by oracle attacks, we forge de-identified faces by superimposing a collection of carefully designed noise patterns onto the original face. The modification of the image is controlled to minimize the probability of good recognition while minimizing the distortion. In addition, these de-identified images are by construction made robust to counter attacks such as blurring. We present an experimental validation in which we de-identify LFW faces and show that resulting images are still recognized by human beings while deceiving a state-of-the-art face recognition algorithm [23].

### *6.5.4. Tardos code in practice*

**Participant:** Teddy Furon.

*Joint work with the Inria supported start-up LAMARK.*

We deal with active fingerprinting a.k.a. traitor tracing where a collusion of dishonest users merges their individual versions of a content to yield a pirated copy. The Tardos codes are one of the most powerful tools to fight against such collusion process by identifying the colluders. Instead of studying as usual the necessary and sufficient code length in a theoretical setup, we adopt the point of view of the practitioner. We call this the *operational mode*, *i.e.* a practical setup where a Tardos code has already been deployed and a pirated copy has been found. This new paradigm shows that the known bounds on the probability of accusing an innocent in the theoretical setup are way too pessimistic. Indeed the practitioner can resort to much tighter bounds because the problem is fundamentally much simpler under the operational mode. In the end, we benchmark under the operational mode several single decoders recently proposed in the literature [32].

## MIMETIC Project-Team

# 6. New Results

## 6.1. Highlights of the Year

### *6.1.1. Link between performance and risk of injury*

**Participants:** Richard Kulpa [contact], Benoit Bideau, Michaël Ropars.

In our previous biomechanical analysis of the tennis serve, we have demonstrated that the energy flow is a pathomechanical factor, that means that it can increase joint constraints (and thus risk of injury) while not increasing performance. Nevertheless, the definition and evaluation of energy flow is still a complex scientific challenge. We have proposed to compare the energy flow during the serve between injured and non-injured tennis players by investigating the relationships between the quality and magnitude of energy flow, the ball velocity and the peaks of upper limb joint kinetics [11]. The results showed that ball velocity increased and upper limb joint kinetics decreased with the quality of energy flow from the trunk to the 'hand+racket'. Injured players showed a lower quality of energy flow through the upper limb kinetic chain, a lower ball velocity and higher rates of energy absorbed by the shoulder, the elbow than non-injured players. These findings imply that an effective energy flow through the kinetic chain by using a proper serve technique is necessary for reducing overuse joint injury risks.

### *6.1.2. ACM SIGGRAPH Course on crowd simulation*

**Participant:** Julien Pettré [contact].

Crowds for entertainment or safety applications purposes are most of the time simulated using microscopic algorithms. In contrast with other types of approaches, microscopic simulators are able to generate continuous and smooth trajectories for individual agents. They are based on models of local interactions between agents. The crowd motion result form the combination of all local motion and interactions. The fact that the resulting crowd motion is emergent makes difficult anticipating the simulation results. Many motion and interaction models have been proposed to design a plethora of simulation algorithms: force-based models, rule-based models, coupled or not with flow-based models, etc. Each type of interaction models will actually result into specific crowd motions as well as agents trajectories. Unfortunately, not all have the desired properties: oscillations, jerky trajectories, residual collisions or deadlocks are often observed in simulations. From this point of view, the course [28] presents the many recent progresses in crowd simulation algorithms since the introduction of velocity-based algorithms , as well as the impact on the level of realism and the visual quality of simulated crowd motions. It also presents the impact on various kind of applications.

### *6.1.3. Immersive basketball playing*

**Participants:** Franck Multon [contact], Alexandra Covaci, Anne-Hélène Olivier.

The paper has received the best paper award of the ACM VRST 2014 Conference in November 2014. This paper addressed the problem of perception of distances in immersive environments when dealing with precision distant tasks, such as basketball free throw. The work has been done in collaboration with University of Brasov in Romania, thanks to the FP7 VISIONAIR infrastructure project. The main results of this work tend to show that third person perspectives enabled subjects to perform the task with similar movements than in real world, compared to first person perspective. Third person perspective consists in placing the camera at a different place from the eye point of view, as in many videogames. On the opposite first person perspective consists in place the camera at the place of the user's eyes in scale 1:1, as if the user was colocalized in the virtual environment. We also demonsrated an adaptation to the task in immersive environments, which is a key information for future development of training methods using VR. We have been invited to submit an extended version of the paper to the IEEE Computer and Graphics journal for 2015.

BEST PAPERS AWARDS :

[21] **ACM Symposium on Virtual Reality Software and Technology VRST**. A. COVACI, A.-H. OLIVIER, F. MULTON.

## 6.2. VR and Sports

**Participants:** Richard Kulpa [contact], Benoit Bideau, Franck Multon.

Previous works in MimeTIC have shown the advantage of using VR to design and carry-out experiments on perception-action coupling in sports, especially for duels between two opponents. However the impact of using various technical solutions to carry-out this type of experiment in sports is not clear. Indeed immersion is performed by using interfaces to capture the motion/intention of the user and to deliver various multi-sensory feedbacks. These interfaces may affect the perception-action loop so that results obtained in VR cannot be systematically transferred to real practice.

Most of the applications in VR provide the user with visual feedbacks in which the avatar of the user can be more or less simplified (sometimes limited to a hand or the tools he his carrying). In first person view in caves the user generally does not need accurate avatars as he can perceive his real body but some authors have shown that the perception of distances is generally modified. Some authors have also demonstrated that first person view was less efficient that third person view with avatars when performing accurate tasks such as reaching objects in constrained environments. We proposed an experiment to evaluate which type of feedback was the most appropriate one for complex precision tasks, such as basketball free-throw. In basketball free-throw the user has to throw a ball into a small basket placed at over 4.5m far from him. Thus perception of distance is actually a key point in such a task. Beginners and experts carried-out a first experiment in real in order to measure their motion and performance in real situation. Then beginners were asked to perform free throws with a real ball in hands, but in three conditions in a Cave (Immersia): 1) first person view (see Figure 4 ), 2) third-person view with the visual feedback of the ball's position, and 3) third-person view the virtual ball and additional rings modeling the perfect trajectory for the ball to get in the basket. Results show that significant difference exist in ball speed between first person view condition compared to real condition whereas no difference exist in third-person view conditions. If we focus on successful throws only, ball speed in the last condition 3) was very similar to real condition whereas all the other VR conditions (1) and 2)) lead to significant differences compared to real situation. In all VR conditions the height of ball release was significantly higher in VR compared to real situation. These results show that VR conditions lead to adaptations in the way people perform such a precision task, especially for ball speed and height of ball release. However this difference is significantly higher with first person view and tends to zero in condition 3). Future works will tend to evaluate new conditions with avatars and complementary points of view (such as lateral and frontal views together as suggested by some authors). It will also be important to more clearly understand the problem of perception of distances in such an environment. This work has been performed in cooperation with University of Brassov in Romania [21]. This paper has received the best paper award of the ACM VRST 2014 Conference.

Another key feedback is the external forces associated with the task. In most sports applications such forces are strongly linked to performance. However delivering these forces in virtual environments is still a challenge as it required haptic devices that could affect the way the users perform the task (with a different grip compared to real situation and limitations in dynamic response of the device). Pseudohaptic has been introduced in the early 2000. It consists in using visual feedbacks to make people perceive the forces linked to a task. However this approach has not been tested for whole-body interaction. In collaboration with Hybrid team in Inria Rennes, we studied how the visual animation of a self-avatar could be artificially modified in real-time in order to generate different haptic perceptions. In our experimental setup participants could watch their self-avatar in a virtual environment in mirror mode. They could map their gestures on the self-animated avatar in real-time using a Kinect. The experimental task consisted in a weight lifting with virtual dumbbells that participants could manipulate by means of a tangible stick. We introduce three kinds of modification of the visual animation of the self-avatar: 1) an amplification (or reduction) of the user motion (change in C/D ratio), 2) a change in the dynamic profile of the motion (temporal animation), or 3) a change in the posture of the avatar (angle of inclination). An example is depicted in Figure 5 . Thus, to simulate the lifting of a "heavy" dumbbell, the

*Figure 4. First-person view condition in the basket free-throw.*

avatar animation was distorted in real-time using: an amplification of the user motion, a slower dynamics, and a larger angle of inclination of the avatar. We evaluated the potential of each technique using an ordering task with four different virtual weights. Our results show that the ordering task could be well achieved with every technique. The C/D ratio-based technique was found the most efficient. But participants globally appreciated all the different visual effects, and best results could be observed in the combination configuration [9]. Our results pave the way to the exploitation of such novel techniques in various VR applications such as for sport training, exercise games, or industrial training scenarios in single or collaborative mode.

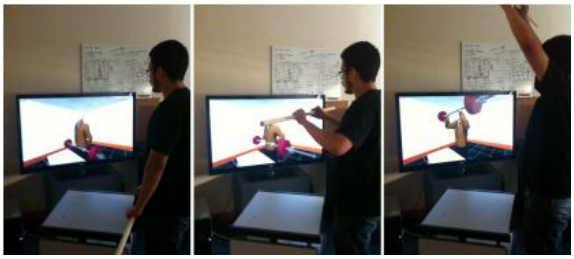

*Figure 5. Weight discrimination task: the animation of the avatar showed a lifting effort according to the weight of the virtual dumbbell and the user has to rank the feebacks from the lightest to the heaviest mass.*

## 6.3. Motion Sensing and analysis

**Participant:** Franck Multon [contact].

Sensing human activity is a very active field of research, with a wide range of applications ranging from entertainment and serious games to personal ambient living assistance, including rehabilitation. MimeTIC aims at proposing original methods to process raw motion capture data in order to compute relevant information according to the application.

In rehabilitation, we have collaborated with University of Montreal, Saint-Justine Hospital which main activity is rehabilitation of children with pathologies of the pyramidal control system. In this domain, defining metrics and relevant measurement to diagnose pathologies and to monitor patients during treatment is a key point. In gait, most of the previous works focus on gait spatio-temporal parameters (such as step length, frequency, stride duration, global speed...) which could be measured with two main families of systems: 1) one-point measurement with a force plate, one accelerometer or dedicated devices (such as a Gait Ride), or 2) multi-point measurement systems with motion sensors or markers placed over the patient's skin. The former provides the clinician with compact but incomplete knowledge whereas the latter provides him with numerous data which are sometimes difficult to analyze and to get (specific technical skills are required). The first step to any type of analysis is to detect the main gait events, such as foot strikes and toe offs. In treadmill walking, widely used in rehabilitation as it enables the clinician to analyze numerous gait cycles in a limited place with a controlled speed, automatically detecting such gait events requires complex devices with specific technical skills (such as calibration and post-processing with motion capture systems) [1].

Recent papers have demonstrated that low-cost and easy-to-use depth cameras (such as a Kinect from Microsoft) look promising for serious applications requiring motion capture. However there exist some confusion between the feet and the ground at foot strike and foot off leading to bad estimation of the gait cycle events. We have proposed an alternative approach that consists in using the strong correlation between knee and foot trajectories to deduce foot strikes thanks to knee movements. Indeed, the maximal distance between knees along the longitudinal axis provides us with very accurate gait events detection compared to previous works. We have validated this detection events on walking patterns that were also altered by placing a 5cm-sole below the left (resp. right) foot of the subject to create asymetry. The results show that this gait cycle event detection based on depth images is as accurate as using reference methods based on accurate motion capture systems.

## 6.4. VR and Ergonomics

**Participants:** Charles Pontonnier [contact], Georges Dumont, Pierre Plantard, Franck Multon.

The use of virtual reality tools for ergonomics applications is a very important challenge in order to genezalize the use of such devices for the design of workstations.

We proposed in collaboration with Thierry Duval (Lab-Stic, Telecom Bretagne, Brest) a new architecture for information sharing and bridging in collaborative virtual environements in application to ergonomics studies. We particularly presented this year how we implemented the design engineer role in the collaborative environment [30], [29]. We are currently evaluating the complete framework for collaborative ergonomics by defining use-cases and trying to find the best design mode to efficiently solve this problem. Moreover, we have developed and evaluated some manipulation techniques, such as the 7-handle technique which is particularly efficient to manipulate large objects in an immersive environments [6], [27]. A demonstration of this technique has been presented during the ICAT-EGVE conference [37].

We also contributed in the on-site motion analysis field. Microsoft Kinect is a promising tool to evaluate human poses without markers, calibration and manual post-processing. It has been applied to a wide set of applications, such as entertainment, rehabilitation, sports analysis and more recently in ergonomics. In MimeTIC we wish to develop innovative approaches based on a Kinect in order to assess the potential risks of musculoskeletal disorders. However analyzing humans in work places is challenging because of many potential occlusions and displacements of the user. Hence it is a key point to evaluate to which extent this method could be applied to real work places, in real work condition. Most of previous works aiming at evaluating the Kinect sensor generally focus on simple 2D poses. In this work we proposed to evaluate the reliability of Kinect measurements for assessing the movement of operators in ergonomic studies with complex 3D upper-limb poses [38]. To this end we asked subjects to perform complex 3D arm motions concurrently measured with a Kinect and a Vicon motion capture system. The results demonstrated that most of the poses were correctly estimated with the Kinect but specific poses are badly reconstructed, leaded to errors going up 30°. Hence, experimenter should take this information into account when using a Kinect on a work place in order to avoid experimenting these bad results.

At last, we proposed a new approach for the use of virtual reality with haptics in the Product Development Process loop for testing deformable parts by introducing the user in the loop and proposing a two-stage deformation simulation method for real time haptic interaction. Such an approach is of importance to let the designer be able to handle and validate the design of a product or a workstation respecting multiple constraints, e.g. ergonomics, bulk or productivity. This approach has been fully detailed in a book chapter published this year [34]

## 6.5. Virtual Human Animation

**Participants:** Franck Multon [contact], Julien Pettré, Steve Tonneau.

A common issue in three-dimensional animation is the creation of contacts between a virtual creature and the environment. Contacts allow force exertion, which produces motion. This paper addresses the problem of computing contact configurations allowing to perform motion tasks such as getting up from a sofa, pushing an object or climbing. We propose a two-step method to generate contact configurations suitable for such tasks. The first step is an offline sampling of the range of motion (ROM) of a virtual creature. The ROM of the human arms and legs is precisely determined experimentally. The second step is a run time request confronting the samples with the current environment. The best contact configurations are then selected according to a heuristic for task efficiency. The heuristic is inspired by the force transmission ratio. Given a contact configuration, it measures the potential force that can be exerted in a given direction. The contact configurations are then used as inputs for an inverse kinematics solver that will compute the final animation. Our method is automatic and does not require examples or motion capture data. It is suitable for real time applications and applies to arbitrary creatures in arbitrary environments. Various scenarios (such as climbing, crawling, getting up, pushing or pulling objects) are used to demonstrate that our method enhances motion autonomy and interactivity in constrained environments [15], [32]. In Figure 6 , a character is able to select the most appropriate constraints to pull a heavy cupboard by putting a foot on an obstacle to maximize the force ratio.

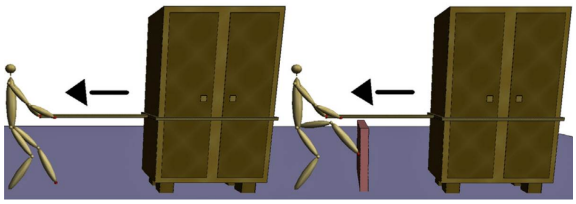

*Figure 6. First-person view condition in the basket free-throw.*

## 6.6. Biomechanics for avatar animation

**Participants:** Charles Pontonnier, Georges Dumont, Steve Tonneau, Franck Multon, Julien Pettré [contact], Ana Lucia Cruz Ruiz, Antoine Muller.

Bio-inpired controllers and planners are compelling for avatar animation. We are currently engaging several works on the subject within the frame of the ENTRACTE project.

Ana-Lucia Cruz-Ruiz has been recruited as a PhD student since november 2013. The goal of this thesis is to define and evaluate muscle-based controllers for avatar animation.A first result has been obtained in defining and validating a bio-inspired limb controller based on a linearizing loop of a neuromuscular complex. Application on a one-dof limb has been validated by comparing the muscle activation shapes obtained in simulation with standard records of biceps and triceps activation [3].

## 6.7. Semantically consistent herarchical decomposition of virtual urban environments

**Participants:** Carl-Johan Jorgensen, Fabrice Lamarche [contact].

When planning a path in their environment, pedestrians do not consider every detail at once. Instead, people first plan a coarse path, choosing streets to travel to reach their goal. Local decisions such as where to cross a street or on which side to pass by a pole are taken during navigation. In computer science, hierarchical representations of an environment are often used to reduce the computation cost of the planning algorithm. Such a representation also enables smarter navigation behaviours. Indeed, it others the opportunity to delay the local planning until relevant information is available. It also enables a quick recovery from unexpected events, as the high-level path might stay valid even if unexpected events alter the lower-level path.

We proposed a method that automatically generates a three-level hierarchical representation of an informed urban environment. In this hierarchy, each level is a semantically coherent partition of the navigation areas and can be used to plan paths at different levels of abstraction. This representation is used in a path planning process that delays some decisions until relevant information is perceived. This algorithm uses path options to smartly adapt the path when unexpected events occur.

## 6.8. Shoulder biomechanics

**Participants:** Armel Crétual [contact], Michaël Ropars.

At first sight, in clinical practice, shoulder mobility is frequently evaluated through mono-axial amplitude. Interestingly, for diagnosing shoulder hyperlaxity or frozen shoulder, external rotation of the arm whilst at the side (ER1) is commonly used. We first gave a definition of hyperlaxity, as described actually in the literature, and its link with shoulder instability and treatment. After looking for an optimized way to examine external rotation of the shoulder, we proposed the definition of a novel index to quantify global shoulder mobility, the Shoulder Configuration Space Volume (SCSV) corresponding to the reachable volume in the configuration space of the shoulder joint [4]. Then, this new index was examined through correlation to shoulder signs of hyperlaxity [14].

**MYRIADS Project-Team**

# 5. New Results

## 5.1. Highlights of the Year

- The Contrail project coordinated by Christine Morin received the "Excellent" grade at its final review held on March 14th, 2014 in Brussels.
- Anne-Cécile Orgerie has been awarded the Young Researcher prize of the Lyon city in November 2014.
- Christine Morin has been awarded one of the 12 "Etoile de l'Europe 2014" prizes in December 2014 for the coordination of the Contrail European project.

BEST PAPERS AWARDS :
[18] **4th International Conference on Cloud Computing and Services Science**. H. FERNANDEZ, C. STRATAN, G. PIERRE.

## 5.2. Dependable Cloud Computing

**Participants:** Jiajun Cao, Stéphane Chevalier, Gene Cooperman, Teodor Crivat, Roberto-Gioacchino Cascella, Stefania Costache, Florian Dudouet, Filippo Gaudenzi, Anna Giannakou, Yvon Jégou, Ancuta Iordache, Christine Morin, Anne-Cécile Orgerie, Edouard Outin, Nikolaos Parlavantzas, Jean-Louis Pazat, Guillaume Pierre, Aboozar Rajabi, Louis Rilling, Matthieu Simonin, Arnab Sinha, Cédric Tedeschi.

### *5.2.1. Deployment of distributed applications in a multi-provider environment*

**Participants:** Roberto-Gioacchino Cascella, Stefania Costache, Florian Dudouet, Filippo Gaudenzi, Yvon Jégou, Christine Morin, Arnab Sinha.

The move of users and organizations to Cloud computing will become possible when they are able to exploit their own applications, applications and services provided by cloud providers, as well as applications from third party providers in a trustful way on different cloud infrastructures. In the framework of the Contrail European project [2] [46], we have designed and implemented the Virtual Execution Platform (VEP) service in charge of managing the whole life cycle of OVF distributed applications under Service Level Agreement rules on different infrastructure providers [47]. In 2013, we designed the CIMI inspired REST-API for VEP 2.0 with support for Constrained Execution Environment (CEE), advance reservation and scheduling service, and support for SLAs [56], [55] [57]. We integrated support for delegated certificates and developed test scripts to integrate the Virtual Infrastructure Network (VIN) service. VEP 1.1 was slightly modified to integrate the usage control (Policy Enforcement Point (PEP)) solution developed by CNR. The CEE management interface was developed during 2013 and is available through the graphical API as well as through the RESTful API.

### *5.2.2. Checkpointing for multi-cloud environments*

**Participants:** Jiajun Cao, Gene Cooperman, Christine Morin, Matthieu Simonin.

Most cloud platforms currently rely on each application to provide its own fault tolerance. A uniform mechanism within the cloud itself serves two purposes: (a) direct support for long-running jobs, which would otherwise require a custom fault-tolerant mechanism for each application; and (b) the administrative capability to manage an over-subscribed cloud by temporarily swapping out jobs when higher priority jobs arrive.

We propose ([31]) a novel *Checkpointing as a Service* approach, which enables application checkpointing and migration in heterogeneous cloud environments. Our approach is based on a non-invasive mechanism to add fault tolerance to an existing cloud platform *after the fact*, with little or no modification to the cloud platform itself. It achieves its cloud-agnostic property by using an external checkpointing package, independent of the target cloud platform. We implemented a prototype of the service on top of both OpenStack and Snooze IaaS clouds. We conducted a preliminary performance evaluation using the Grid'5000 experimentation platform.

### 5.2.3. Towards a distributed cloud inside the backbone

**Participants:** Anne-Cécile Orgerie, Cédric Tedeschi.

The DISCOVERY proposal currently in phase of construction and lead by Adrien Lèbre from the ASCOLA team, and currently on leave at Inria aims at designing a distributed cloud, leveraging the resources we can find in the network's backbone. [0]

In this context, and in collaboration with ASCOLA and ASAP teams, we started the design of an overlay network whose purpose is to be able, with a limited cost, to locate geographically-close nodes from any point of the network. The design, implementation, and experimentation of the overlay has been described in an article published in 2014 [22].

### 5.2.4. A multi-objective adaptation system for the management of a Distributed Cloud

**Participants:** Yvon Jégou, Edouard Outin, Jean-Louis Pazat.

In this project, we consider a "Distributed Cloud" made of multiple data/computing centers interconnected by a high speed network. A distributed Cloud is neither a usual Cloud built around a single data center, nor a Cloud Federation interconnecting different data centers owned and run by different administrative entities. Moreover, in the Cloud organization targeted here, the network capabilities can be dynamically configured in order to apply optimizations to guarantee QoS for streaming or negotiated bandwidth for example. Due to the dynamic capabilities of the Clouds, often referred to as elasticity, there is a strong need to dynamically adapt both platforms and applications to users needs and environmental constraints such as electrical power consumption.

We address the management of the Distributed Cloud in order to consider both optimizations for energy consumption and for users' QoS needs. The objectives of these optimizations will be negotiated as contracts on Service Level Agreement (SLA). A special emphasis will be put on the distributed aspect of the platform and include both servers and network adaptation capabilities. The design of the system will rely on self-* techniques and on adaptation mechanisms at any level (from IaaS to SaaS). The MAPE-k framework (Monitor-Analysis-Planning-Execution based on knowledge) will be used for the implementation of the system. The technical developments are based on the Openstack framework.

This work is done in cooperation with the DIVERSE team and in cooperation with Orange under the umbrella of the B-COM Technology Research Center.

### 5.2.5. Multi-cloud application deployment in ConPaaS

**Participants:** Stéphane Chevalier, Teodor Crivat, Guillaume Pierre.

We extended ConPaaS to support the deployment of smartphone backend applications in mobile operators' base stations. The motivation is to reduce the latency compared to a traditional deployment where the backend is located in an external cloud. This requires building a lightweight infrastructure which allows one to easily create containers that can be seamlessly migrated (roaming). A publication on this topic will appear in 2015 [23].

### 5.2.6. Application Performance Modeling in Heterogeneous Cloud Environments

**Participants:** Ancuta Iordache, Guillaume Pierre.

Heterogeneous cloud platforms offer many possibilities for applications for make fine-grained choice over the types of resources they execute on. This opens for example opportunities for fine-grained control of the tradeoff between expensive resources likely to deliver high levels of performance, and slower resources likely to cost less. We designed a methodology for automatically exploring this performance vs. cost tradeoff when an arbitrary application is submitted to the platform. Thereafter, the system can automatically select the set of resources which is likely to implement the tradeoff specified by the user. We significantly improved the speed at which the system can characterize the performance of an arbitrary application. A publication on this topic is currently under review.

[0] The DISCOVERY website: http://beyondtheclouds.github.io

### 5.2.7. Dynamic reconfiguration for multi-cloud applications

**Participants:** Nikolaos Parlavantzas, Aboozar Rajabi.

In the context of the PaaSage European project, we are working on model-based self-optimisation of multi-cloud applications. In particular, we are developing a dynamic adaptation system, capable of transforming the currently running application configuration into a target configuration in a cost-effective and safe manner. In 2014, we have defined the architecture of the adaptation system and produced a first prototype[30].

### 5.2.8. Self-adaptable Monitoring for Security in the Cloud

**Participants:** Anna Giannakou, Christine Morin, Jean-Louis Pazat, Louis Rilling.

We aim at designing a self-adaptable system for security monitoring in clouds. The considered system should cope with the dynamic nature of virtual infrastructures in clouds and have a minimal impact on performance. In 2014, we studied the state of the art in cloud security monitoring, which is composed of various approaches for intrusion detection systems (IDS), based on traditional IDS techniques such as signature-based detection and anomaly-based detection.

As a first step towards our goal of making self-adaptable a complete security monitoring architecture for cloud environments, we defined a simple initial monitoring scenario for identifying the impact of the dynamicity of a cloud architecture on the intrusion detection process. In this scenario, the security monitoring infrastructure is composed of two network IDS instances, which are used to monitor the virtual infrastructures network traffic of two cloud clients (one virtual infrastructure per client), and also eventually monitor the physical infrastructure (that is the operator's infrastructure). The virtual network traffic in each host machine is monitored by only one of the IDS instances, so that the IDS instances must be adapted to topology changes (such as migration of VMs) in the cloud environment. The adaptation process includes updates of the rules configured in the instance (deletion or creation).

In 2014, we built our testbed based on OpenStack technology for the underlying IaaS cloud platform and Snort for the network IDS. At this point the testbed consists of only five machines (on the Grid'5000 platform) but we aim to increase the number of host machines and deploy more VMs for having a more realistic representation of a production network. This will allow us to study performance issues and also more complex security monitoring setups. Our goal is also to enable monitoring of other elements, such as resource usage (both per host and per VM) on the cloud provider side.

### 5.2.9. Fog Computing

**Participant:** Jean-Louis Pazat.

The concept of "Fog Computing" is currently developed on the idea of hosting instances of services, not on centralized datacenters (i.e. the "Cloud"), but on a highly distributed infrastructure: the Internet Edge (i.e. the "Fog"). This infrastructure consists in geographically distributed computing resources with relatively small capabilities. Compared with datacenters, a "Fog" infrastructure is able to offer to Service Providers a shorter distance from the service to the user but with the same flexibility of software deployment and management.

This work focus on the problem of resource allocation in such infrastructure when considering services in the area of Internet of Things, Social Networks or Online Gaming. For such use-cases, service-to-user latency is a critical parameter for the quality of experience. Optimizing such parameter is an objective for the platform built on top of the Fog Infrastructure that will be dedicated to the deployment of the considered service. In order to achieve such a goal, the platform needs to select some strategies for the allocation of network and computing resources, based on the initial requirements for the service distribution.

We first focus on the formal expression of these requirements, by considering first the requirements provided by a Service Operator to the "Fog" Infrastructure (required computing resources, minimal quality of experience (QoE) level, etc.). The resource allocation strategies should also take into account the topology of the "Fog" Infrastructure, the heterogeneous capabilities of the equipments and of the underlying network. Based on this information, strategies and algorithms for resource allocation should be designed that will participate in the process of building an efficient platform for the service distribution. Evaluation of this efficiency will be an important process to justify the relevance of the strategies.

This work is part of Bruno Stevant's PhD thesis that began in December 2014. It is done in cooperation with the REOP team, Institut Mines telecom/IRISA.

## 5.3. Heterogeneous Resource Management

**Participants:** Eliya Buyukkaya, Djawida Dib, Eugen Feller, Christine Morin, Nikolaos Parlavantzas, Guillaume Pierre.

### *5.3.1. Cross-resource scheduling in heterogeneous cloud environments*

**Participants:** Eliya Buyukkaya, Guillaume Pierre.

Allocating resources to applications in a heterogeneous cloud environment is harder than in a homogeneous environment. In a heterogeneous cloud some rare resources are more precious than others, and should be treated carefully to maximize their utilization. Similarly, applications may request groups of resources that exhibit certain inter-resource properties such as the available bandwidth between the assigned resources. We are currently investigating scheduling algorithms for handling such scenarios.

### *5.3.2. Maximizing private cloud provider profit in cloud bursting scenarios*

**Participants:** Djawida Dib, Christine Morin, Nikolaos Parlavantzas.

Current PaaS offerings either provide no support for SLA guarantees or provide limited support targeting a restricted set of application types. To overcome this limitation, we have developed an open, cloud-bursting PaaS system, called Meryn, designed to be easily extensible to host new application types. The system integrates a decentralized optimization policy that maximises the PaaS provider profit, taking into account the payment of penalties incurred when quality guarantees are unsatisfied. The system was implemented and evaluated on the Grid5000 testbed using batch and MapReduce workloads. The results demonstrated the effectiveness of the policy in increasing provider profit [16] This work was part of Djawida Dib's PhD thesis [10] defended in July 2014.

### *5.3.3. Data life-cycle management in clouds*

**Participants:** Eugen Feller, Christine Morin.

Infrastructure as a Service (IaaS) clouds provide a flexible environment where users can choose and control various aspects of the machines of interest. However, the flexibility of IaaS clouds presents unique challenges for storage and data management in these environments. Users use manual and/or ad-hoc methods to manage storage and data in these environments. FRIEDA is a Flexible Robust Intelligent Elastic Data Management framework that employs a range of data management strategies approaches in elastic environments. This year, our work carried out in the context of the DALHIS associate team [0], was focused on the extended design and evaluation of the FRIEDA data management system. FRIEDA was tested to work on Amazon EC2 resources. In addition, we layered a commandline utility atop FRIEDA that allows users to plug-in applications to run in FRIEDA. These tools have been adopted by the LBL-ATLAS group to run their experiments on Amazon [29].

## 5.4. Energy-efficient Resource Infrastructures

**Participants:** Maria Del Mar Callau Zori, Alexandra Carpen-Amarie, Bogdan Florin Cornea, Ismael Cuadrado Cordero, Djawida Dib, Eugen Feller, Sabbir Hasan Rochi, Yunbo Li, Christine Morin, Anne-Cécile Orgerie, Jean-Louis Pazat, Guillaume Pierre, Lavinia Samoila.

### *5.4.1. Energy-efficient IaaS clouds*

**Participants:** Alexandra Carpen-Amarie, Christine Morin, Anne-Cécile Orgerie.

[0] http://project.inria.fr/dalhis

Energy consumption has always been a major concern in the design and cost of data centers. The wide adoption of virtualization and cloud computing has added another layer of complexity to enabling an energy-efficient use of computing power in large-scale settings. Among the many aspects that influence the energy consumption of a cloud system, the hardware-component level is one of the most intensively studied. However, higher-level factors such as virtual machine properties, their placement policies or application workloads may play an essential role in defining the power consumption profile of a given cloud system. In this work, we explored the energy consumption patterns of Infrastructure-as-a-Service (IaaS) cloud environments under various synthetic and real application workloads. For each scenario, we investigated the power overhead triggered by different types of virtual machines, the impact of the virtual cluster size on the energy-efficiency of the hosting infrastructure and the tradeoff between performance and energy consumption of MapReduce virtual clusters through typical cloud applications [45].

### 5.4.2. Energy-aware IaaS-PaaS co-design

**Participants:** Maria Del Mar Callau Zori, Alexandra Carpen-Amarie, Djawida Dib, Anne-Cécile Orgerie, Guillaume Pierre, Lavinia Samoila.

The wide adoption of the cloud computing paradigm plays a crucial role in the ever-increasing demand for energy-efficient data centers. Driven by this requirement, cloud providers resort to a variety of techniques to improve energy usage at each level of the cloud computing stack. However, prior studies mostly consider resource-level energy optimizations in IaaS clouds, overlooking the workload-related information locked at higher levels, such as PaaS clouds. We argue that cross-layer cooperation in clouds is a key to achieving an optimized resource management, both performance and energy-wise. To this end, we claim there is a need for a cooperation API between IaaS and PaaS clouds, enabling each layer to share specific information and to trigger correlated decisions. We identified the drawbacks raised by such co-design objectives and discuss opportunities for energy usage optimizations. A position paper has been published on these aspects [15]. Ongoing work is currently conducted in order to quantify the actual possible gains both energy and performance-wise for this IaaS-PaaS co-design approach.

### 5.4.3. Energy-efficient and network-aware resource allocation in Cloud infrastructures

**Participants:** Ismael Cuadrado Cordero, Christine Morin, Anne-Cécile Orgerie.

Cloud computing is increasingly becoming an essential component for Internet service provision, yet at the same time its energy consumption has become a key environmental and economic concern. It becomes urgent to improve the energy efficiency of such infrastructures. Our work aims at designing energy-efficient resource allocation for Cloud infrastructures. Yet, energy is not the only criterion to take into account at risk of losing users. A multi-criteria approach is required in this context to satisfy both users and Cloud providers.

The proposed resource allocation algorithms will take into account not only the computing resources but also the storage and networking resources. Indeed, the ever-growing appetite of new applications for network resources leads to an unprecedented electricity bill for network resources, and for these bandwidth-hungry applications, networks can become an significant bottleneck. This phenomenon is emphasized with the emergence of the big data paradigm. The designed algorithms would thus integrate the data locality dimension to optimize computing resource allocation while taking into account the fluctuating limits of network resources.

In 2014, several experiments were performed to understand and quantify networking energy consumption. These experiments include network protocol energy consumption in the devices, configuration energy consumption in switching/routing devices and associated energy consumption to real cloud computing applications (e.g. Google drive). These experiments have been performed over systems provided by Inria such as Grid'5000 and specific network devices (e.g. level 3 router for a private LAN). Based on this work, we developed an analytic model of networking energy consumption in a cloud computing environment. This analysis will serve as a basis for designing an energy-efficient architecture and related algorithms.

### 5.4.4. Simulating Energy Consumption of Wired Networks

**Participants:** Bogdan Florin Cornea, Anne-Cécile Orgerie.

Predicting the performance of applications, in terms of completion time and resource usage for instance, is critical to appropriately dimension resources that will be allocated to these applications. Current applications, such as web servers and Cloud services, require lots of computing and networking resources. Yet, these resource demands are highly fluctuating over time. Thus, adequately and dynamically dimension these resources is challenging and crucial to guarantee performance and cost-effectiveness. In the same manner, estimating the energy consumption of applications deployed over heterogeneous cloud resources is important in order to provision power resources and make use of renewable energies. Concerning the consumption of entire infrastructures, some studies show that computing resources represent the biggest part in Cloud's consumption, while others show that, depending on the studied scenario, the energy cost of the network infrastructure that links the user to the computing resources can be bigger than the energy cost of the servers. In this work, we aim at simulating the energy consumption of wired networks which receive little attention in the Cloud computing community even though they represent key elements of these distributed architectures. To this end, we are contributing to the well-known open-source simulator ns3 by developing an energy consumption module named ECOFEN. Through this tool, we have studied the energy consumption of data transfers in Clouds [19]. This work has been done in collaboration with the Avalon team from LIP in Lyon.

### *5.4.5. Resource allocation in a Cloud partially powered by renewable energy sources*

**Participants:** Yunbo Li, Anne-Cécile Orgerie.

We propose here to design a disruptive approach to Cloud resource management which takes advantage of renewable energy availability to perform opportunistic tasks. To begin with, the considered Cloud is mono-site (i.e. all resources are in the same physical location) and performs tasks (like web hosting or MapReduce tasks) running in virtual machines. This Cloud receives a fixed amount of power from the regular electric Grid. This power allows it to run usual tasks. In addition, this Cloud is also connected to renewable energy sources (such as windmills or solar cells) and when these sources produce electricity, the Cloud can use it to run more tasks.

The proposed resource management system needs to integrate a prediction model to be able to forecast these extra-power periods of time in order to schedule more work during these periods. Batteries will be used to guarantee that enough energy is available when switching on a new server working exclusively on renewable energy. Given a reliable prediction model, it is possible to design a scheduling algorithm that aims at optimizing resource utilization and energy usage, problem known to be NP-hard. The proposed heuristics will thus schedule tasks spatially (on the appropriate servers) and temporally (over time, with tasks that can be planed in the future).

This work is done in collaboration with Ascola team from LINA in Nantes.

### *5.4.6. SLA driven Cloud Auto-scaling for optimizing energy footprint*

**Participants:** Sabbir Hasan Rochi, Jean-Louis Pazat.

As a direct consequence of the increasing popularity of Internet and Cloud Computing services, data centers are amazingly growing and hence have to urgently face energy consumption issues. At the Infrastructure-as-a-Service (IaaS) layer, Cloud Computing allows to dynamically adjust the provision of physical resources according to Platform-as-a-Service (PaaS) needs while optimizing energy efficiency of the data center.

The management of elastic resources in Clouds according to fluctuating workloads in the Software-as-a-Service (SaaS) applications and different Quality-of-Service (QoS) end-user's expectations is a complex issue and cannot be done dynamically by a human intervention. We advocate the adoption of Autonomic Computing (AC) at each XaaS layer for responsiveness and autonomy in front of environment changes. At the SaaS layer, AC enables applications to react to a highly variable workload by dynamically adjusting the amount of resources in order to keep the QoS for the end users. Similarly, at the IaaS layer, AC enables the infrastructure to react to context changes by optimizing the allocation of resources and thereby reduce the costs related to energy consumption. However, problems may occur since those self-managed systems are related in some way (e.g. applications depend on services provided by a cloud infrastructure): decisions taken in isolation at given layer may interfere with other layers, leading whole system to undesired states.

We propose an approach driven by Service Level Agreements (SLAs) for Cloud auto-scaling. A SLA defines a formal contract between a service provider and a service consumer on an expected QoS level. The main idea of this thesis is to exploit the SLA requirements to (i) avoid the interferences between the Cloud autonomic managers by a cross-layer coordination of SLA contracts; (ii) fine-tune the resources needs according to SLA by proposing both dynamic resources provisioning for optimizing the energy footprint and dynamic reconfiguration at the SaaS level to optimize the expected QoS. In particular, we propose to address renewable energy in the SLA contract. The objective is twofold. First, for ecological reasons, it allows Cloud users to express their preferences about the energy provider and the nature of the energy in the data center. Then, for economic reasons, it takes advantage of renewable energy costs (expressed in the SLA) to reconfigure resource allocation and energy usage. The integration of such SLAs in each layer of the Cloud stack and their management by an autonomic manager or by the coordination of autonomic managers still remain open issues.

This work is done in collaboration with Ascola team from LINA in Nantes.

### *5.4.7. Simulating the impact of DVFS within SimGrid*

**Participants:** Alexandra Carpen-Amarie, Christine Morin, Anne-Cécile Orgerie.

Simulation is a a popular approach for studying the performance of HPC applications in a variety of scenarios. However, simulators do not typically provide insights on the energy consumption of the simulated platforms. Furthermore, studying the impact of application configuration choices on energy is a difficult task, as not many platforms are equipped with the proper power measurement tools. The goal of this work is to enable energy-aware experimentations within the SimGrid simulation toolkit, by introducing a model of application energy consumption and enabling the use of Dynamic Voltage and Frequency Scaling (DVFS) techniques for the simulated platforms. We provide the methodology used to obtain accurate energy estimations, highlighting the simulator calibration phase. The proposed energy model is validated by means of a large set of experiments featuring several benchmarks and scientific applications. This work is available in the latest SimGrid release. This work is done in collaboration with the Mescal team from LIG in Grenoble.

## 5.5. Decentralised and Adaptive workflows

**Participants:** Christine Morin, Jean-Louis Pazat, Javier Rojas Balderrama, Matthieu Simonin, Cédric Tedeschi, Palakyiem Wallah.

### *5.5.1. Template workflows*

**Participants:** Christine Morin, Javier Rojas Balderrama, Matthieu Simonin, Cédric Tedeschi.

In the framework of the DALHIS associate team [0], we started to combine the high-level template workflow language TIGRES [0], developed by our partner team from Lawrence Berkeley National Lab (LBL) with the workflow management system developed in the team [5]. The design of this integration and its benefits have been presented in a workhsop article [24].

### *5.5.2. Adaptive Workflows with Chemical Computing*

**Participants:** Javier Rojas Balderrama, Matthieu Simonin, Cédric Tedeschi.

We are currently designing a complete programming model for the management of adaptive workflows, based on an extension of the HOCL language, in particular workflows that may evolve at run time in their shape. An article is under preparation.

### *5.5.3. Best-effort decentralised workflow execution*

**Participants:** Jean-Louis Pazat, Cédric Tedeschi, Palakyiem Wallah.

[0] http://project.inria.fr/dalhis
[0] http://tigres.lbl.gov/home

We are currently proposing a simple workflow model for workflow execution in platforms with limited computing resources and services. The key idea is to devise a best-effort workflow engine that does not require a strong centralised orchestrator. Such a workflow engine relies on point-to-point cooperation between nodes supporting the execution.

## 5.6. Experimental Platforms

**Participants:** Maxence Dunnewind, Nicolas Lebreton, Julien Lefeuvre, David Margery, Eric Poupart.

### *5.6.1. Energy measurement*

**Participants:** Maxence Dunnewind, Nicolas Lebreton, David Margery, Eric Poupart.

In the context of the $ECO_2$Clouds project, the BonFIRE infrastructure was updated. At the software layer, the complete monitoring stack was revisited so as to attribute power consumption values to all VMs running on the infrastructure and to expose this information to users. This was used by the project partners to confirm that using an eco-aware scheduler could significantly reduce eco-impact of running a distributed infrastructure.

### *5.6.2. BonFIRE*

**Participants:** Maxence Dunnewind, Julien Lefeuvre, David Margery, Eric Poupart.

The project was reviewed in December 2013 during CloudCom 2013 in Bristol and rated Excellent. It has been kept in working state through our commitment to the BonFIRE foundation. The main acheivment on this topic was to evolve the cloud reservation system so as to support tracking usage using allocation blocks, as a fragment of the physical machines. Instance types can therefore have a different footprint in number of allocation blocks depending on the hardware they are scheduled on.

### *5.6.3. Fed4FIRE*

**Participants:** Nicolas Lebreton, Julien Lefeuvre, David Margery.

In Fed4FIRE, two key technologies have been adopted as common protocols to enable experimenters to interact with testbeds: Slice Federation Architecture (SFA), to provision resources, and Control and Management Framework for Networking Testbeds (OMF) to control them. Here, we contributed to a proposal to secure usage of OMF and to a design to allow using BonFIRE through SFA. In 2014, the main area of work has been maintenance of the infrastructure and initial prototyping of an SFA API to BonFIRE.

# PANAMA Project-Team

# 6. New Results

## 6.1. Highlights of the Year

The EUSIPCO 2014 Best Student Paper Award was awarded to our joint paper [32] on dynamic screening for sparse regularization. .

A review paper on audio source separation, rooted in METISS/PANAMA know-how and contributions to this topic over the years, was published in the IEEE Signal Processing Magazine [25].

A new version of the Flexible Audio Source Separation Toolbox (FASST) was released in January 2014 and downloaded 300 times.

BEST PAPERS AWARDS :

[32] **European Signal Processing Conference EUSIPCO 2014**. A. BONNEFOY, V. EMIYA, L. RALAIVOLA, R. GRIBONVAL.

## 6.2. Recent results on sparse representations

Sparse approximation, high dimension, scalable algorithms, dictionary design, sample complexity

The team has had a substantial activity ranging from theoretical results to algorithmic design and software contributions in the field of sparse representations, which is at the core of the ERC project PLEASE (projections, Learning and Sparsity for Efficient Data Processing, see Section 8.2.1.1 ).

### *6.2.1. A new framework for sparse representations: analysis sparse models*

**Participants:** Rémi Gribonval, Nancy Bertin, Srdan Kitic, Cagdas Bilen, Laurent Albera.

In the past decade there has been a great interest in a synthesis-based model for signals, based on sparse and redundant representations. Such a model assumes that the signal of interest can be composed as a linear combination of *few* columns from a given matrix (the dictionary). An alternative *analysis-based* model can be envisioned, where an analysis operator multiplies the signal, leading to a *cosparse* outcome. Within the SMALL FET-Open project, we initiated a research programme dedicated to this analysis model, in the context of a generic missing data problem (e.g., compressed sensing, inpainting, source separation, etc.). We obtained a uniqueness result for the solution of this problem, based on properties of the analysis operator and the measurement matrix. We also considered a number of pursuit algorithms for solving the missing data problem, including an $\ell^1$-based and a new greedy method called GAP (Greedy Analysis Pursuit). Our simulations demonstrated the appeal of the analysis model, and the success of the pursuit techniques presented.

These results have been published in conferences and in a journal paper [100]. Other algorithms based on iterative cosparse projections [83] as well as extensions of GAP to deal with noise and structure in the cosparse representation have been developed, with applications to toy MRI reconstruction problems and acoustic source localization and reconstruction from few measurements [101].

Successful applications of the cosparse approach to sound source localization, audio declipping and brain imaging have been developed this year. In particular, we compared the performance of several cosparse recovery algorithms in the context of sound source localization [97] and showed its efficiency in situations where usual methods fail ([37], see paragraph 6.6.3 ). It was also shown to be applicable to the hard declipping problem [49]. Application to EEG brain imaging was also investigated and a paper was published in MLSP14 [28] (see paragraph 6.6.4 ).

### *6.2.2. Theoretical results on sparse representations*

**Participants:** Rémi Gribonval, Anthony Bourrier, Pierre Machart, Yann Traonmilin, Gilles Puy.

*Main collaboration: Mike Davies (University of Edinburgh), Patrick Perez (Technicolor R&I France), Tomer Peleg (The Technion)*

**Fundamental performance limits for ideal decoders in high-dimensional linear inverse problems:** The primary challenge in linear inverse problems is to design stable and robust "decoders" to reconstruct high-dimensional vectors from a low-dimensional observation through a linear operator. Sparsity, low-rank, and related assumptions are typically exploited to design decoders which performance is then bounded based on some measure of deviation from the idealized model, typically using a norm. We characterized the fundamental performance limits that can be expected from an ideal decoder given a general model, ie, a general subset of "simple" vectors of interest. First, we extended the so-called notion of instance optimality of a decoder to settings where one only wishes to reconstruct some part of the original high dimensional vector from a low-dimensional observation. This covers practical settings such as medical imaging of a region of interest, or audio source separation when one is only interested in estimating the contribution of a specific instrument to a musical recording. We defined instance optimality relatively to a model much beyond the traditional framework of sparse recovery, and characterized the existence of an instance optimal decoder in terms of joint properties of the model and the considered linear operator [106], [105]. This year, noiseless and noise-robust settings were both considered in the journal paper [16]. We showed somewhat surprisingly that the existence of noise-aware instance optimal decoders for all noise levels implies the existence of a noise-blind decoder. A consequence of our results is that for models that are rich enough to contain an orthonormal basis, the existence of an L2/L2 instance optimal decoder is only possible when the linear operator is not substantially dimension-reducing. This covers well-known cases (sparse vectors, low-rank matrices) as well as a number of seemingly new situations (structured sparsity and sparse inverse covariance matrices for instance). We exhibit an operator-dependent norm which, under a model-specific generalization of the Restricted Isometry Property (RIP), always yields a feasible instance optimality and implies instance optimality with certain familiar atomic norms such as the $\ell^1$ norm. Current work explores the existence of convex decoders for general union of subspaces models under generalized RIP assumptions, as well as conditions ensuring that random low-dimensional projections ensure the RIP even when the projection is from an infinite-dimensional space to a finite dimensional one. Envisioned applications are in compressive learning (see Section 6.4 ).

**Connections between sparse approximation and Bayesian estimation:** Penalized least squares regression is often used for signal denoising and inverse problems, and is commonly interpreted in a Bayesian framework as a Maximum A Posteriori (MAP) estimator, the penalty function being the negative logarithm of the prior. For example, the widely used quadratic program (with an $\ell^1$ penalty) associated to the LASSO / Basis Pursuit Denoising is very often considered as MAP estimation under a Laplacian prior in the context of additive white Gaussian noise (AWGN) reduction.

In 2011 we obtained a result [85] highlighting the fact that, while this is *one* possible Bayesian interpretation, there can be other equally acceptable Bayesian interpretations. Therefore, solving a penalized least squares regression problem with penalty $\phi(x)$ need not be interpreted as assuming a prior $C \cdot \exp(-\phi(x))$ and using the MAP estimator. In particular, we showed that for *any* prior $P_X$, the minimum mean square error (MMSE) estimator is the solution of a penalized least square problem with some penalty $\phi(x)$, which can be interpreted as the MAP estimator with the prior $C \cdot \exp(-\phi(x))$. Vice-versa, for *certain* penalties $\phi(x)$, the solution of the penalized least squares problem is indeed the MMSE estimator, with a certain prior $P_X$. In general $dP_X(x) \neq C \cdot \exp\left(-\phi(x)\right)dx$. In 2013, we extended this result to general inverse problems [88], [86], [87]. This year, we worked on the characterization of such relations beyond the Gaussian noise model, with the objective of understanding whether similar results hold when the quadratic data-fidelity term is replaced with other convex losses.

### 6.2.3. Algorithmic and theoretical results on dictionary learning

**Participants:** Rémi Gribonval, Nancy Bertin, Srdan Kitic, Cagdas Bilen, Luc Le Magoarou, Melanie Ducoffe.

*Main collaboration (theory for dictionary learning): Rodolphe Jenatton, Francis Bach (Equipe-projet SIERRA (Inria, Paris)), Martin Kleinsteuber, Matthias Seibert (TU-Munich),*

*Main collaboration (dictionary learning for gesture recognition): Anatole Lecuyer, Ferran Argelaguet (EPI HYBRID, Rennes)*

**Theoretical guarantees for dictionary learning :** 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 litterature to learn a dictionary from a collection of training samples, there are little existing results which provide an adequate mathematical understanding of the behaviour of these techniques and their ability to recover an ideal dictionary from which the training samples may have been generated.

Beyond our pioneering work [89], [109] [6] on this topic, which concentrated on the noiseless case for non-overcomplete dictionaries, this year we obtained new results showing the relevance of an $\ell^1$ penalized cost function for the locally stable identification of overcomplete incoherent dictionaries, in the presence of noise and outliers [54]. Moreover, we established new sample complexity bounds of dictionary learning and other related matrix factorization schemes (including PCA, NMF, structured sparsity ...) [55], [46], [38].

**Learning computationally efficient dictionaries** Classical dictionary learning is limited to small-scale problems. Inspired by usual fast transforms, we proposed a general dictionary structure that allows cheaper manipulation, and an algorithm to learn such dictionaries –and their fast implementation [50]. A preprint is available [56], a paper will appear at ICASSP 2015, and a journal paper is in preparation.

**Operator learning for cosparse representations :** Besides standard dictionary learning, we also considered learning in the context of the cosparse model. The overall problem is to learn a low-dimensional signal model from a collection of training samples. The mainstream approach is to learn an overcomplete dictionary to provide good approximations of the training samples using sparse synthesis coefficients. This famous sparse model has a less well known counterpart, in analysis form, called the cosparse analysis model. In this new model, signals are characterized by their parsimony in a transformed domain using an overcomplete analysis operator.

In specific situations, when prior information is available on the operator, it is possible to express it in parametric form and learn this parameter. For instance, in the sound source localization problem, we showed that the unknown speed of sound can be learned jointly in the process of cosparse recovery, under mild conditions. This work was presented at iTwist'14 workshop [48].

**Dictionary learning for gesture modeling** In collaboration with the HYBRID project-team (internship of Melanie Ducoffe), we explored the potential of dictionary learning in the context of motion tracking. Motion tracking technology, especially for commodity hardware, requires robust gesture recognition algorithms to fully exploit the benefits of natural user interfaces. We proposed a gesture recognition algorithm based on the sparse representation of motion data, with a learning phase consisting in learning a dictionary of basic gestures. A paper is in preparation.

## 6.3. Activities on waveform design for telecommunications

**Participant:** Rémi Gribonval.

*Main collaboration: Marwa Chafii, Jacques Palicot, Carlos Bader (Equipe SCEE, Supelec, Rennes)*

Peak to Average Power Ratio (PAPR), Orthogonal Frequency Division Multiplexing (OFDM), Generalized Waveforms for Multi Carrier (GWMC)

In the context of the TEPN (Towards Energy Proportional Networks) Comin Labs project (see Section 8.1.1.2 ), in collaboration with the SCEE team at Supelec (thesis of Marwa Chafii co-supervised by R. Gribonval), we investigated a problem related to dictionary design: the characterization of waveforms with low Peak to Average Power Ratio (PAPR) for wireless communications. This is motivated by the importance of a low PAPR for energy-efficient transmission systems. A first stage of the work consisted in characterizing the statistical distribution of the PAPR for a general family of multi-carrier systems, leading to a journal paper [17] and several conference communications [27], [33]. The work now concentrates on characterizing waveforms with optimum PAPR.

## 6.4. Emerging activities on compressive sensing, learning and inverse problems

Compressive sensing, compressive learning, acoustic wavefields, audio inpainting,

### *6.4.1. Audio inpainting*

**Participants:** Rémi Gribonval, Nancy Bertin, Corentin Guichaoua, Srdan Kitic, Anh Tho Le.

Inpainting is a particular kind of inverse problems that has been extensively addressed in the recent years in the field of image processing. It consists in reconstructing a set of missing pixels in an image based on the observation of the remaining pixels. Sparse representations have proved to be particularly appropriate to address this problem. However, inpainting audio data has never been defined as such so far. A series of works about audio inpainting was initiated by the METISS team in the framework of the EU Framework 7 FET-Open project FP7-ICT-225913-SMALL (Sparse Models, Algorithms and Learning for Large-Scale data).

Building upon our previous contributions (definition of the audio inpainting problem as a general framework for many audio processing tasks, application to the audio declipping or desaturation problem, formulation as a sparse recovery problem [60]), new results were obtained this year to address the case of audio declipping with the competitive cosparse approach. Its promising results, especially when the clipping level is low, were confirmed experimentally by the formulation and use of a new algorithm named Cosparse Iterative Hard Tresholding, which is a counterpart of the sparse Consistent Iterative Hard Thresholding. These results were presented during the iTwist'14 workshop [49]. Additional experiments were performed (internship of Anh Tho Le) to confirm the results on a larger database and investigate optimal parameters (nature and redundancy of the dictionary, relaxation parameter for the cosparsity level).

Current and future works deal with developing advanced (co)sparse decomposition for audio inpainting, including several forms of structured sparsity (*e.g.* temporal and multichannel joint-sparsity), dictionary learning for inpainting, and several applicative scenarios (declipping, time-frequency inpainting).

### *6.4.2. Blind Calibration of Compressive Sensing systems*

**Participants:** Rémi Gribonval, Cagdas Bilen, Gilles Puy.

*Main collaborations: Gilles Chardon, Laurent Daudet (Institut Langevin)*

We consider the problem of calibrating a compressed sensing measurement system under the assumption that the decalibration consists of unknown gains on each measure. We focus on blind calibration, using measures performed on a few unknown (but sparse) signals. A naive formulation of this blind calibration problem, using $\ell^1$ minimization, is reminiscent of blind source separation and dictionary learning, which are known to be highly non-convex and riddled with local minima. In the considered context, when the gains are real valued and non-negative, we showed that in fact this formulation can be exactly expressed as a convex optimization problem, and can be solved using off-the-shelf algorithms. Numerical simulations demonstrated the effectiveness of the approach even for highly uncalibrated measures, when a sufficient number of (unknown, but sparse) calibrating signals is provided. We observed that the success/failure of the approach seems to obey sharp phase transitions [84]. We extended the framework to phase-only decalibration, using techniques revolving around low-rank matrix recovery [66], [65], [110], [64], and to joint phase and gain decalibration [15].

### *6.4.3. Compressive Gaussian Mixture estimation*

**Participants:** Rémi Gribonval, Anthony Bourrier, Nicolas Keriven.

*Main collaborations: Patrick Perez (Technicolor R&I France)*

When fitting a probability model to voluminous data, memory and computational time can become prohibitive. In this paper, we propose a framework aimed at fitting a mixture of isotropic Gaussians to data vectors by computing a low-dimensional sketch of the data. The sketch represents empirical moments of the underlying probability distribution. Deriving a reconstruction algorithm by analogy with compressive sensing, we experimentally show that it is possible to precisely estimate the mixture parameters provided that the sketch is large enough. Our algorithm provides good reconstruction and scales to higher dimensions than previous probability

mixture estimation algorithms, while consuming less memory in the case of numerous data. It also provides a privacy-preserving data analysis tool, since the sketch does not disclose information about individual datum it is based on [71], [69], [70]. This year, extensions to non-isotropic Gaussians, with new algorithms and preliminary applications to speaker verification have been conducted.

## 6.5. Recent results on tensor decompositions

Multi-linear algebra is defined as the algebra of $q$-way arrays ($q > 2$), that is, the arrays whose elements are addressed by more than two indices. The first works back as far as Jordan who was interested in simultaneously diagonalizing two matrices at a time [96]. It is noteworthy that such two matrices can be interpreted as both slices of a three-way array and their joint diagonalization can be viewed as Hitchcock's polyadic decomposition [92] of the associated three-way array. Other works followed discussing rank problems related to multi-way structures and properties of multi-way arrays. However, these exercices in multilinear algebra were not linked to real data analysis but stayed within the realm of mathematics. Studying three-way data really started with Tucker's seminal work, which gave birth to the three-mode factor analysis [114]. His model is now often referred to as the Tucker3 model. At the same moment, other authors focused on a particular case of the Tucker3 model, calling it PARAFAC for PARAllel FACtor analysis [91], and on the means to achieve such a decomposition, which will become the famous canonical decomposition [74]. In honor to Hitchcock's pionneer work, we will call it the Canonical Polyadic (CP) decomposition.

Achieving a CP decomposition has been seen first as a mere non-linear least squares problem, with a simple objective criterion. In fact, the objective is a polynomial function of many variables, where some separate. One could think that this kind of objective is easy because smooth, and even infinitely differentiable. But it turns out that things are much more complicated than they may appear to be at first glance. Nevertheless, the Alternating Least Squares (ALS) algorithm has been mostly utilized to address this minimization problem, because of its programming simplicity. This should not hide the inherently complicated theory that lies behind the optimization problem. Moreover, in most of the applications, actual tensors may not exactly satisfy the expected model, so that the problem is eventually an approximation rather than an exact decomposition. This may results in a slow convergence (or lack of convergence) of iterative algorithms such as the ALS one [98]. Consequently, a new class of efficient algorithms able to take into account the properties of tensors to be decomposed is needed.

### *6.5.1. CP decomposition of semi-symmetric semi-nonnegative three-way arrays*

**Participant:** Laurent Albera.

*Main collaboration (Line search and trust region strategies): Julie Coloigner (LTSI, France), Amar Kachenoura (LTSI, France), Lotfi Senhadji (LTSI, France)*

*Main collaborations (Jacobi-like approaches): Lu Wang (LTSI, France), Amar Kachenoura (LTSI, France), Lotfi Senhadji (LTSI, France), Huazhong Shu (LIST, China)*

We proposed new algorithms for the CP decomposition of semi-nonnegative semi-symmetric three-way tensors. In fact, it consists in fitting the CP model for which two of the three loading matrices are nonnegative and equal. Note that such a problem can also be interpreted as a nonnegative Joint Diagonalization by Congruence (JDC) problem.

**Line search and trust region strategies**

We first circumvented the nonnegativity constraint by means of changes of variable into squares, leading to a (polynomial) unconstrained optimization problem. Two optimization strategies, namely line search and trust region, were then studied. Regarding the former, a global plane search scheme was considered. It consists in computing, for a given direction, one or two optimal stepsizes, depending on whether the same stepsize is used in various updating rules. Moreover, we provided a compact matrix form for the derivatives of the objective function. This allows for a direct implementation of several iterative algorithms such as Conjugate Gradient (CG), Levenberg-Marquardt (LM) and Newton-like methods, in matrix programming environments like MATLAB. Note that the computational complexity issue was taken into account in the design phase of the algorithms, and was evaluated for each algorithm, allowing to fairly compare their performance.

Thus, various scenarios have been considered, aiming at testing the influence of i) an additive noise, which can stand for modeling errors, ii) the collinearity between factors, iii) the array rank and iv) the data size. The comparisons between our CG-like, Newton-like and LM-like methods (where semi-nonnegativity and semi-symmetry constraints are exploited), and classical CP algorithms (where no constraints are considered), showed that a better CP decomposition is obtained when these a priori are exploited, especially in the context of high dimensions and high collinearity. Finally, based on our numerical analysis, the algorithms that seem to yield the best tradeoff between accuracy and complexity are our $CG_{2steps}$-like and LM-like algorithms. This work was published in the Elsevier Linear Algebra and Applications journal [19].

Next, we considered an exponential change of variable leading to a different (non-polynomial) unconstrained optimization problem. Then we proposed novel algorithms based on line search strategy with an analytic global plane search procedure requiring new matrix derivations. Their performance was evaluated in terms of estimation accuracy and computational complexity. The classical ELS-ALS [108] and LM [112] algorithms without symmetry and nonnegativity constraints, and the ACDC algorithm [115] where only the semi-symmetry constraint is imposed, were tested as reference methods. Furthermore, the performance was also compared with our algorithms based on a square change of variable. The comparison studies showed that, among these approaches, the best accuracy/complexity trade off was achieved when an exponential change of variable was used through our ELS-ALS-like algorithm. This work was published in the Elsevier Signal Processing journal [18].

**Jacobi-like approaches**

The line search (despite the use of global plane search procedures) and trust region strategies may be sensitive to initialization, and generally require a multi-initialization procedure. In order to circumvent this drawback, we considered in this work Jacobi-like approaches, which are known to be less sensitive to initialization. Note that our line search and trust region approaches can then be used to refine the solution obtained by the latter.

More particularly, we formulated the high-dimensional optimization problem into several sequential polynomial and rational subproblems using i) a square change of variables to impose nonnegativity and ii) LU or QR matrix factorization for parameterization. The two equal nonnegative loading matrices are actually written as the Hadamard product of two equal matrices which can be factorized as the product of elementary matrices, each one depending on only one parameter.

The proposed approach reduces the optimization problem to the computation of the two equal nonnegative loading matrices only. The third loading matrix is algebraically derived from the latter. This requires an appropriate parameterization of the set of matrices whose inverse is nonnegative. This work was published in a journal paper [26]. Numerical experiments on simulated matrices emphasize the advantages of the proposed algorithms over classical CP and JDC techniques, especially in the case of degeneracies.

## 6.6. Source separation and localization

Source separation, sparse representations, tensor decompositions, semi-nonnegative independent component analysis, probabilistic model, source localization

### *6.6.1. A general framework for audio source separation*

**Participants:** Frédéric Bimbot, Rémi Gribonval, Nancy Bertin.

*Main collaboration: E. Vincent, Y. Salaün (EPI PAROLE, Inria Nancy); A. Ozerov, N.Q.K. Duong (Technicolor R&I France)*

Source separation is the task of retrieving the source signals underlying a multichannel mixture signal.

About a decade ago, state-of-the-art approaches consisted of representing the signals in the time-frequency domain and estimating the source coefficients by sparse decomposition in that basis. These approaches rely only on spatial cues, which are often not sufficient to discriminate the sources unambiguously. Over the last years, we proposed a general probabilistic framework for the joint exploitation of spatial and spectral cues [102], which generalizes a number of existing techniques including our former study on spectral GMMs [61]. 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., and it became the basis of a large number of works in the field, including our own. In the last years, improvements were obtained through the use of prior knowledge about the source spatial covariance matrices [81], [95], [94], knowledge on the source positions and room characteristics [82], or a better initialization of parameters thanks to specific source localization techniques [68].

This accumulated progress lead to two main achievements this year, which show the maturity of our work and which will leverage its impact. First, a new version of the Flexible Audio Source Separation Toolbox, fully reimplemented, was released. It will provide the community with an efficient and ergonomic software, making available the tools from past years' research [58]. Second, we published an overview paper on recent and going research along the path of *guided* separation, *i.e.*, techniques and models allowing to incorporate knowledge in the process towards efficient and robust solutions to the audio source separation problem, in a special issue of IEEE Signal Processing Magazine devoted to source separation and its applications [25].

### 6.6.2. Towards real-world separation and remixing applications

**Participants:** Nancy Bertin, Frédéric Bimbot, Jules Espiau de Lamaestre, Anaik Olivero, Jérémy Paret, Nathan Souviraà -Labastie.

*Emmanuel Vincent (EPI PAROLE, Inria Nancy)*

While some challenges remain, work from previous years and our review paper on guided source separation [25] highlighted that progress has been made and that audio source separation is closer than ever to successful industrial applications, especially when some knowledge can be incorporated. This is exemplified by the contract with MAIA Studio, which reaches its end in December 2014 and showed in particular how user input or side information could raise source separation tools to efficient solutions in real-world applications.

In this context, new tools were developed this year. The introduction of manually-tuned parameters in the automated separation process, which modifies the Wiener filtering coefficients obtained from estimation of the mixtures covariance matrices, allows to find a better trade-off between artifacts and interferences. In order to ensure high audio quality for such applications, some user-guided corrections remain necessary even after an automatic pre-separation ; to this end, we developed an improved display (based on cepstrum and automatic constrast adaptation) and semi-automatic selection and suppression tools in the time-frequency domain. Those tools take as few inputs as possible from the user and their result can be ergonomically adjusted from the baseline output to a manually fine-tuned area, in a very small operating time. We also proposed tools to suppress a time-frequency area and replace it by content extracted from its context, reducing the perceptual impact of the suppression.

In some applicative contexts of source separation, several mixtures are available which contain similar instances of a given source. We have designed a general framework for audio source separation guided by multiple audio references, where each audio reference is a mixture which is supposed to contain at least one source similar to one of the target sources. Deformations between the sources of interest and their references are modeled in a general manner. A nonnegative *matrix co-factorization* algorithm is used which allows sharing of information between the considered mixtures. We have experimented our algorithm on music plus voice mixtures with music and/or voice references. Applied on movies and TV series data, the algorithm improves the signal-to-distortion ratio (SDR) of the sources of lowest intensity by 9 to 12 decibels with respect to original mixture [40]

### 6.6.3. Acoustic source localization

**Participant:** Nancy Bertin, Srdan Kitic, Laurent Albera, Nancy Bertin, Rémi Gribonval.

*Main collaborations (audio-based control for robotics): Aly Magassouba and François Chaumette (Inria, EPI LAGADIC, France)*

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. Common techniques, including beamforming, only provides the *direction of arrival* of the sound, estimated from the *Time Difference of Arrival* (*TDOA*) [68]. This year, we have particularly investigated alternative approaches, either where the explicit localization is not needed (audio-based control of a robot) or, on the contrary, where the exact location of the source is needed and/or TDOA is irrelevant (cosparse modeling of the acoustic field).

**Implicit localization through audio-based control for robotics**

In robotics, the use of aural perception has received recently a growing interest but still remains marginal in comparison to vision. Yet audio sensing is a valid alternative or complement to vision in robotics, for instance in homing tasks. Most existing works are based on the relative localization of a defined system with respect to a sound source, and the control scheme is generally designed separately from the localization system.

In contrast, the approach that we started investigating this year focuses on a sensor-based control approach. We proposed a new line of work, by considering the hearing sense as a direct and real-time input of closed loop control scheme for a robotic task. Thus, and unlike most previous works, this approach does not necessitate any explicit source localization: instead of solving the localization problem, we focus on developing an innovative modeling based on sound features. To address this objective, we placed ourselves in the sensor-based control framework, especially visual servoing (VS) that has been widely studied in the past [76].

From now on, we have established an analytical model linking sound features and control input of the robot, defined and analyzed robotic homing tasks involving multiple sound sources, and validated the proposed approach by simulations. This work is mainly lead by Aly Magassouba, whose Ph.D. is co-supervised by Nancy Bertin and François Chaumette. A conference paper presenting these first results was submitted to ICRA 2015. Future work will include real-world experiments with the robot Romeo from Aldebaran Robotics.

**Cosparse modeling of the acoustic field**

Cosparse modeling is particularly attractive when the signals of interest satisfy certain physical laws that naturally drive the choice of an analysis operator, which is the case for the acoustic field, ruled by the wave equation. Unlike usual localization techniques such as beamforming or TDOA-based direction estimation, which generally consider reverberation as an adverse condition, the cosparse modeling of sound propagation has also the interest of considering reverberation as a source of additional information for the localization task. Eventually, it can provide a full coordinate localization of the sources, and not only their direction of arrival.

Building upon our previous results on cosparse modeling and recovery algorithms for the wave equation [97], we have obtained additional evidence of the interest of this approach. In particular, we have showed that recasting source localization as a cosparse inverse problem allows to scale up to 3-dimensional problems which were untractable with the counterpart sparse approach. Moreover, we have confirmed that our model takes indeed advantage of reverberation, by showing that localization remains possible when the sources and the microphones are partly separated by an acoustically opaque obstacle (a situation where TDOA would obviously fail). These two results were published and presented during ICASSP'14 [37]. Recent results also include algorithmic improvements (through the use of the Alternating Direction Method of Multipliers (ADMM) framework), and evidence that, in addition to its scaling capabilities, the sparse analysis computational cost can even *benefit* from an increase in the number of measurements. A journal paper including these new results and presenting them jointly with co-space modeling in the context of brain source localization (see Section 6.6.4 ) is under preparation.

### *6.6.4. Brain source localization*

**Participants:** Laurent Albera, Srdan Kitic, Nancy Bertin, Rémi Gribonval.

*Main collaborations (tensor-based approaches): Hanna Becker (GIPSA & LTSI, France), Isabelle Merlet (LTSI, France), Fabrice Wendling (LTSI, France), Pierre Comon (GIPSA, France), Christian Benar (La Timone, Marseille), Martine Gavaret (La Timone, Marseille), Gwénaël Birot (FBML, Genève), Martin Haardt (TUI, Germany)*

*Main collaborations (from tensor to sparse models): Hanna Becker (GIPSA & LTSI, France), Pierre Comon (GIPSA, France), Isabelle Merlet (LTSI, France), Fabrice Wendling (LTSI, France)*

*Main collaborations (a sparsity-based approach): Hanna Becker (Technicolor, France), Pierre Comon (GIPSA, France), Isabelle Merlet (LTSI, France)*

*Main collaborations (a multimodal sparsity-based approach): Thomas Oberlin, Pierre Maurel, Christian Barillot (EPI VISAGES, Rennes, France)*

**Tensor-based approaches**

The localization of several simultaneously active brain regions having low signal-to-noise ratios is a difficult task. To do this, tensor-based preprocessing can be applied, which consists in constructing a Space-Time-Frequency (STF) or Space-Time-Wave-Vector (STWV) tensor and decomposing it using the CP decomposition. We proposed a new algorithm for the accurate localization of extended sources based on the results of the tensor decomposition. Furthermore, we conducted a detailed study of the tensor-based preprocessing methods, including an analysis of their theoretical foundation, their computational complexity, and their performance for realistic simulated data in comparison to three conventional source localization algorithms, namely sLORETA [104], cortical LORETA (cLORETA) [103], and 4-ExSo-MUSIC [67]. Our objective consisted, on the one hand, in demonstrating the gain in performance that can be achieved by tensor-based preprocessing, and, on the other hand, in pointing out the limits and drawbacks of this method. Finally, we validated the STF and STWV techniques on real epileptic measurements to demonstrate their usefulness for practical applications. This work was published in the Elsevier NeuroImage journal [13].

**From tensor to sparse models**

The brain source imaging problem has been widely studied during the last decades, giving rise to an impressive number of methods using different priors. Nevertheless, a thorough study of the latter, including especially sparse and tensor-based approaches, is still missing. Consequently, we proposed i) a taxonomy of the methods based on a priori assumptions, ii) a detailed description of representative algorithms, iii) a review of identifiability results and convergence properties of different techniques, and iv) a performance comparison of the selected methods on identical data sets. Our aim was to provide a reference study in the biomedical engineering domain which may also be of interest for other areas such as wireless communications, audio source localization, and image processing where ill-posed linear inverse problems are encountered and to identify promising directions for future research in this area. A part of this work was presented at ICASSP'14 [30] while the whole part was submitted to IEEE Signal Processing Magazine.

**A cosparsity-based approach**

Cosparse modeling is particularly attractive when the signals of interest satisfy certain physical laws that naturally drive the choice of an analysis operator. We showed how to derive a reduced non-singular analysis operator describing EEG signals from Poisson's equation, Kirchhoff's law and some other physical constraints. As a result, we proposed the CoRE (Cosparse Representation of EEG signals) method to solve the classical brain source imaging problem. Computer simulations demonstrated the numerical performance of the CoRE method in comparison to a dictionary-based sparse approach. This work was partially presented at MLSP'14 [28].

**A sparsity-based approach**

Identifying the location and spatial extent of several highly correlated and simultaneously active brain sources from EEG recordings and extracting the corresponding brain signals is a challenging problem. In our comparison of source imaging techniques presented at ICASSP'14 [30], the VB-SCCD algorithm [79], which exploits the sparsity of the variational map of the sources, proved to be a promising approach. We proposed several ways to improve this method. In order to adjust the size of the estimated sources, we added a regularization term that imposes sparsity in the original source domain. Furthermore, we demonstrated the application of ADMM, which permitted to efficiently solve the optimization problem. Finally, we also considered the exploitation of the temporal structure of the data by employing L1,2-norm regularization. The performance of the resulting algorithm, called L1,2-SVB-SCCD, was evaluated based on realistic simulations in comparison to VB-SCCD and several state-of-the-art techniques for extended source localization. This work was partially presented at EUSIPCO'14 [29] and a journal paper is in preparation.

**A multimodal sparsity-based approach**

In the context of the HEMISFER Comin Labs project (see Section 8.1.1.1 ), in collaboration with the VISAGES team, we investigated brain imaging using simultaneously recorded electroencephalography (EEG) and functional magnetic resonance imaging (fMRI). To this end, we introduced a linear coupling model that links the electrical EEG signal to the hemodynamic response from the blood-oxygen level dependent (BOLD) signal. Both modalities are then symmetrically integrated, to achieve a high resolution in time and space while allowing some robustness against potential decoupling of the BOLD effect. The joint imaging problem is expressed as a linear inverse problem, which is addressed using sparse regularization. The sparsity prior naturally reflects the fact that only few areas of the brain are activated at a certain time, and it is easily implemented through proximal algorithms. At this stage, the signifiance of the method and its effectiveness have been demonstrated through numerical investigations on a simplified head model and simulated data on a realistic brain model. A conference paper has been submitted and a journal paper is in preparation.

### 6.6.5. *Independent component analysis*

**Participant:** Laurent Albera.

*Main collaboration: Sepideh Hajipour (LTSI & BiSIPL), Isabelle Merlet (LTSI, France), Mohammad Bagher Shamsollahi (BiSIPL, Iran)*

Independent Component Analysis (ICA) is a very useful tool to process biomedical signals including EEG data. We proposed a Jacobi-like Deflationary ICA algorithm, named JDICA. More particularly, while a projection-based deflation scheme inspired by Delfosse and Loubaton's ICA technique (DelL$^{\mathbb{R}}$) [78] was used, a Jacobi-like optimization strategy was proposed in order to maximize a fourth order cumulant-based contrast built from whitened observations. Experimental results obtained from simulated epileptic data mixed with a real muscular activity and from the comparison in terms of performance and numerical complexity with the FastICA [93], RobustICA [116] and DelL$^{\mathbb{R}}$ algorithms, show that the proposed algorithm offers the best trade-off between performance and numerical complexity. This work was submitted for publication in the IEEE Signal Processing Letters journal.

### 6.6.6. *Semi-nonnegative independent component analysis*

**Participant:** Laurent Albera.

*Main collaboration: Lu Wang (LTSI, France), Amar Kachenoura (LTSI, France), Lotfi Senhadji (LTSI, France), Huazhong Shu (LIST, China)*

ICA plays also an important role in many other areas including speech and audio [62], [63], [75], [72], radiocommunications [77] and document restoration [113] to cite a few.

For instance in [113], the authors use ICA to restore digital document images in order to improve the text legibility. Indeed, under the statistical independence assumption, authors succeed in separating foreground text and bleed-through/show-through in palimpsest images. Furthermore, authors in [80] use ICA to solve the ambiguity in X-ray images due to multi-object overlappings. They presented a novel object decomposition technique based on multi-energy plane radiographs. This technique selectively enhances an object that is characterized by a specific chemical composition ratio of basis materials while suppressing the other

overlapping objects. Besides, in the context of classification of tissues and more particularly of brain tumors [107], ICA is very effective. In fact, it allows for feature extraction from Magnetic Resonance Spectroscopy (MRS) signals, representing them as a linear combination of tissue spectra, which are as independent as possible [111]. Moreover, using the JADE algorithm [73] applied to a mixture of sound waves computed by means of the constant-Q transform (Fourier transform with log-frequency) of a temporal waveform broken up into a set of time segments, the authors of [72] describe trills as a set of note pairs described by their spectra and corresponding time envelopes. In this case, pitch and timing of each note present in the trill can be easily deduced.

All the aforementioned applications show the high efficiency of the ICA and its robustness to the presence of noise. Despite this high efficiency in resolving the proposed applicative problems, authors did not fully exploit properties enjoyed by the mixing matrix such as its nonnegativity. For instance in [80], the thickness of each organ, which stands for the mixing coefficient, is real positive. Furthermore, reflectance indices in [113] for the background, the overwriting and the underwriting, which correspond to the mixing coefficients, are also nonnegative. Regarding tissue classification from MRS data, each observation is a linear combination of independent spectra with positive weights representing concentrations [90]; the mixing matrix is again nonnegative.

By imposing the nonnegativity of the mixing matrix within the ICA process, we shown through computer results that the extraction quality can be improved. Exploiting the nonnegativity property of the mixing matrix during the ICA process gives rise to what we call semi-nonnegative ICA. More particularly, we performed the latter by computing a constrained joint CP decomposition of cumulant arrays of different orders [99] having the nonnegative mixing matrix as loading matrices. After merging the entries of the cumulant arrays in the same third order array, the reformulated problem follows the semi-symmetric semi-nonnegative CP model defined in section 6.5.1 . Hence we use the new methods described in section 6.5.1 to perform semi-nonnegative ICA. Performance results in audio and biomedical engineering were given in the different papers cited in section 6.5.1 .

## 6.7. Audio and speech content processing

Audio segmentation, speech recognition, motif discovery, audio mining

### *6.7.1. Audio motif discovery*

**Participants:** Frédéric Bimbot, Nathan Souviraà -Labastie.

*This work was performed in close collaboration with Guillaume Gravier from the Limkmedia project-team.*

As an alternative to supervised approaches for multimedia content analysis, where predefined concepts are searched for in the data, we investigate content discovery approaches where knowledge emerge from the data. Following this general philosophy, we pursued work on motif discovery in audio contents.

Audio motif discovery is the task of finding out, without any prior knowledge, all pieces of signals that repeat, eventually allowing variability. The developed algorithms allows discovering and collecting occurrences of repeating patterns in the absence of prior acoustic and linguistic knowledge, or training material.

We have designed a system to create audio thumbnails of spoken content, i.e., short audio summaries representative of the entire content, without resorting to a lexical representation. As an alternative to searching for relevant words and phrases in a transcript, unsupervised motif discovery is here used to find short, word-like, repeating fragments at the signal level without acoustic models. The output of the word discovery algorithm is exploited via a maximum motif coverage criterion to generate a thumbnail in an extractive manner. A limited number of relevant segments are chosen within the data so as to include the maximum number of motifs while remaining short enough and intelligible.

Evaluation has been performed on broadcast news reports with a panel of human listeners judging the quality of the thumbnails. Results indicate that motif-based thumbnails stand btween random thumbnails and ASR-based keywords, however still far behind thumbnails and keywords humanly authored [35].

### *6.7.2. Mobile device for the assistance of users in potentially dangerous situations*

**Participants:** Romain Lebarbenchon, Ewen Camberlein, Frédéric Bimbot.

The S-Pod project is a cooperative project between industry and academia aiming at the development of mobile systems for the detection of potentially dangerous situations in the immediate environment of a user, without requiring his/her active intervention.

In this context, the PANAMA research group is involved in the design of algorithms for the analysis and monitoring of the acoustic scene around the user, yielding information which can be fused with other sources of information (physiological, contextual, etc...) in order to trigger an alarm when needed and subsequent appropriate measures.

This ongoing work is focused on the development of robust techniques for audio scene analysis, including statistical classification of audio segments into threat vs non-threat categories, and the use of spatial information to determine the location of the user with respect to the potential threat.

## 6.8. Music Content Processing and Music Information Retrieval

Acoustic modeling, non-negative matrix factorisation, music language modeling, music structure

### *6.8.1. Music structure representation and decomposition*

**Participants:** Frédéric Bimbot, Corentin Guichaoua, Anaik Olivero.

*Main collaboration: E. Vincent (EPI PAROLE, Inria Nancy), E. Deruty (external consultant)*

Interest has been steadily growing in semantic audio and music information retrieval for the description of music structure, i.e. the global organization of music pieces in terms of large-scale structural units. Our group has defined a detailed methodology for the semiotic description of music structure, based on concepts and criteria which are formulated as generically as possible, i.e. without resorting to intrinsic properties of the musical content, but rather on relationships between musical elements resulting in well-identifiable patterns. The essential principles and practices developed during an annotation effort deployed by our research group on audio data, in the context of the Quaero project, has led to the public release of over 380 annotations of pop songs from three different data sets (http://musicdata.gforge.inria.fr/structureAnnotation.html) documented by a technical report which includes a few case studies and a concise statistical overview of the annotated data [31]. From the algorithmic point of view, we are currently exploring tree-based representation of music structure where a sequence of musical elements and their relationships are modeled hierarchically as the derivation of a context-free grammar. Parsimony criteria and specific cost functions adapted to music patterns are used to learn the grammar rules and the possibility of distorting the rules is introduced to account for variability accross different repetitions of musical segments.

SAGE Project-Team

# 6. New Results

## 6.1. Highlights of the Year

Lionel Lenôtre and his co-authors revisited in a very efficient way the Hastings-Metropolis Algorithm on Markov Chains for Small-Probability Estimation.

## 6.2. Numerical algorithms

### *6.2.1. Hybrid algebraic sparse linear solvers*

**Participants:** Jocelyne Erhel, David Imberti.

Grants and projects: EXA2CT 7.3.1 , C2S@EXA 7.2.3

Publications: [17]

Abstract: Sparse linear systems arise in computational science and engineering. The goal is to reduce the memory requirements and the computational cost, by means of high performance computing algorithms. Krylov methods combined with Domain Decomposition are very efficient. Numerical results show the benefits of our methodology.

### *6.2.2. GMRES and Polynomial Equivalence*

**Participant:** David Imberti.

Grants and projects: EXA2CT 7.3.1 , C2S@EXA 7.2.3

Publications: in preparation.

Abstract: We have established a theoretical link between GMRES and the much simpler problem of polynomial evaluation along with some algebraic structures to describe the most important elements of the GMRES algorithm. We use these structures to show the connection between sequential GMRES and Horner's Rule, s-step GMRES and Dorn's rule, and predict future possible GMRES-like algorithms.

### *6.2.3. Variables s-step GMRES*

**Participant:** David Imberti.

Grants and projects: EXA2CT 7.3.1 , C2S@EXA 7.2.3

Publications: in preparation.

Abstract: We introduce a new variation on s-step GMRES in order to improve its stability, reduce the number of iterations necessary to ensure convergence, and thereby improve parallel performance. In doing so, we develop a new block variant that allows us to express the stability difficulties in s-step GMRES more fully. We use the algebraic structures previous established via the polynomial equivalence to support an intuitive choice for the variation in the s-step procedure, and reinforce its utility in some communication cost estimates.

### *6.2.4. FGMRES dynamics*

**Participant:** David Imberti.

Grants and projects: EXA2CT 7.3.1 , C2S@EXA 7.2.3

Publications: in preparation.

Abstract: The FGMRES algorithm has met with varying success and we detail theoretical relationships between FGMRES and GMRES including a geometric mean conjecture. Further, we build on the current literature regarding GMRES convergence with an analysis of the dynamical properties of FGMRES.

### 6.2.5. RPM Coupling Factors

**Participant:** David Imberti.

Grants and projects: EXA2CT 7.3.1 , C2S@EXA 7.2.3

Publications: in preparation.

Abstract: We have improved the Recursive Projection Method (RPM) with a subspace version that effectively utilizes parallelism. Furthermore, we include a discussion, numerical experiments, and suggestions for the heretofor neglected coupling factor in RPM, and how they influence convergence of the algorithm.

### 6.2.6. Hastings-Metropolis Algorithm on Markov Chains for Small-Probability Estimation

**Participant:** Lionel Lenôtre.

Grants: H2MNO4 7.2.1

Publications: [13]

Abstract: Shielding studies in neutron transport, with Monte Carlo codes, yield challenging problems of small-probability estimation. The particularity of these studies is that the small probability to estimate is formulated in terms of the distribution of a Markov chain, instead of that of a random vector in more classical cases. Thus, it is not straightforward to adapt classical statistical methods, for estimating small probabilities involving random vectors, to these neutron-transport problems. A recent interacting-particle method for small-probability estimation, relying on the Hastings-Metropolis algorithm, is presented. It is shown how to adapt the Hastings-Metropolis algorithm when dealing with Markov chains. A convergence result is also shown. Then, the practical implementation of the resulting method for small-probability estimation is treated in details, for a Monte Carlo shielding study. Finally, it is shown, for this study, that the proposed interacting-particle method considerably outperforms a simple Monte Carlo method, when the probability to estimate is small.

### 6.2.7. A Strategy for the Parallel Implementations of Stochastic Lagrangian Methods

**Participant:** Lionel Lenôtre.

Grants: H2MNO4 7.2.1

Software: PALMTREE 5.3.1

Publications: [34]

Abstract: We present some investigations on the parallelization of a stochastic Lagrangian simulation. For the self sufficiency of this work, we start by recalling the stochastic methods used to solve Parabolic Partial Differential Equations with a few physical remarks. Then, we exhibit different object-oriented ideas for such methods. In order to clearly illustrate these ideas, we give an overview of the library PALMTREE that we developed. After these considerations, we discuss the importance of the management of random numbers and argue for the choice of a particular strategy. To support our point, we show some numerical experiments of this approach, and display a speedup curve of PALMTREE. Then, we discuss the problem in managing the parallelization scheme. Finally, we analyze the parallelization of hybrid simulation for a system of Partial Differential Equations. We use some works done in hydrogeology to demonstrate the power of such a concept to avoid numerical diffusion in the solution of Fokker-Planck Equations and investigate the problem of parallelizing scheme under the constraint entailed by domain decomposition. We conclude with a presentation of the latest design that was created for PALMTREE and give a sketch of the possible work to get a powerful parallelized scheme.

## 6.3. Numerical models and simulations applied to physics

### 6.3.1. Small scale modeling of porous media

**Participants:** Édouard Canot, Salwa Mansour.

Grants: ARPHYMAT 7.4.3 , 7.4.4

Software: GLiMuH 5.4.3

Publications: [22]

Abstract: This study is devoted to the heat transfer between two spherical grains separated by a small gap; dry air is located around the grains and a liquid water meniscus is supposed to be present between them. This problem can be seen as a micro-scale cell of an assembly of solid grains, for which we are looking for the effective thermal conductivity. For a fixed contact angle and according to the volume of the liquid meniscus, two different shapes are possible for the meniscus, giving a "contacting" state (when the liquid makes a true bridge between the two spheres) and a "non-contacting" one (when the liquid is split in two different drops, separated by a thin air layer); the transition between these two states occurs at different times when increasing or decreasing the liquid volume, thus leading to a hysteresis behavior when computing the thermal flux across the domain.

### *6.3.2. Heat and mass transfer modeling in porous media*

**Participants:** Édouard Canot, Salwa Mansour.

Grants: HYDRINV 7.4.5

Software: HeMaTiS (5.4.1 )

Abstract: The physical model of the HeMaTiS code (5.4.1 ) has been recently improved by adding the diffusion process of dry air through the water steam which is created by the evaporation of the water inside the porous medium. In this fashion, not only can the heating stage of the surface of the soil be simulated but also the cooling stage. The application concerns the study of archaeological fires which were used many times a day; the possibility of alternation of heating and cooling may lead to a better interpretation of residual marks left in the ground. Work is in progress to validate the numerical results.

### *6.3.3. Inverse problem for determining the thermo-physical properties of a porous media*

**Participants:** Édouard Canot, Salwa Mansour.

Grants: HYDRINV 7.4.5

Software: TPIP (5.4.2 )

Publications: [23]

Abstract: This study concerns the inverse problem which consists of the estimation of thermophysical properties of the soil knowing the temperature at few selected points of the domain. In order to solve this inverse problem, we used the least square criterion where we try to minimize the error function between real measures and simulated ones. The coupled system composed of the energy equation together with the three sensitivity boundary initial problems resulting from differentiating the basic energy equation with respect to the soil properties must be solved. To overcome the stiffness of our problem (due to the use of Apparent Heat Capacity method), the high nonlinearity of the coupled system and the problem of large residuals we used the Damped Gauss Newton and Levenberg-Marquardt methods. Moreover, we emphasized on the importance of the choice of $\Delta T$ (temperature range over which the phase change occurs) where for a certain initial guess the inverse problem fails to converge. We overcome this problem by chaining the inverse problems using different values of $\Delta T$ and parameters' set.

### *6.3.4. Geodesy*

**Participant:** Bernard Philippe.

Grants: LIRIMA-EPIC 7.4.2 .

Publications: [12].

Abstract: We solve a geodetic inverse problem for the determination of a distribution of point masses (characterized by their intensities and positions), such that the potential generated by them best approximates a given potential field.

## 6.4. Models and simulations for flow and transport in porous media

### *6.4.1. Simulating Diffusion Processes in Discontinuous Media: Benchmark Tests*

**Participant:** Géraldine Pichot.

Grants: H2MN04 7.2.1

Software: SBM 5.2.2

Publications: [33]

Abstract: We present several benchmark tests for Monte Carlo methods for simulating diffusion in one-dimensional discontinuous media, such as the ones arising the geophysics and many other domains. These benchmarks tests are developed according to their physical, statistical, analytic and numerical relevance. We then perform a systematic study on four numerical methods.

### *6.4.2. Uncertainty Quantification and High Performance Computing for flow and transport in porous media*

**Participants:** Jean-Raynald de Dreuzy, Jocelyne Erhel.

Grants: HYDRINV 7.4.5 , H2MN04 7.2.1

Publications: [18]

Abstract: Stochastic models use random fields to represent heterogeneous porous media. Quantities of interest such as macro dispersion are then analyzed from a statistical point of view. In order to get asymptotic values, large scale simulations must be performed, using High Performance Computing. Non-intrusive methods are well-suited for two-level parallelism. Indeed, for each simulation, parallelism is based on domain decomposition for generating the random input and the flow matrix, parallel linear solvers and parallel particle tracker. Also, several simulations, corresponding to randomly drawn samples, can run in parallel. The balance between these two levels depends on the resources available. The software PARADIS implements flow and transport with random data and computation of macro dispersion. Simulations run on supercomputers with large 3D domains.

### *6.4.3. Computation of macro spreading in 3D porous media with uncertain data*

**Participants:** Jean-Raynald de Dreuzy, Jocelyne Erhel, Mestapha Oumouni.

Grants: HYDRINV 7.4.5 , H2MN04 7.2.1

Publications: [15]

Abstract: We consider an heterogeneous porous media where the conductivity is described by probability laws. Thus the velocity, which is solution of the flow equation, is also a random field, taken as input in the transport equation of a solute. The objective is to get statistics about the spreading and the macro dispersion of the solute. We use a mixed finite element method to compute the velocity and a lagrangian method to compute the spreading. Uncertainty is dealt with a classical Monte-Carlo method, which is well-suited for high heterogeneities and small correlation lengths. We give an explicit formulation of the macro dispersion and a priori error estimates. Numerical experiments with large 3D domains are done with the software PARADIS of the platform H2OLab.

### *6.4.4. A combined collocation and Monte-Carlo method for advection-diffusion equation of a solute in random porous media*

**Participants:** Jocelyne Erhel, Mestapha Oumouni.

Grants: HYDRINV 7.4.5 , H2MN04 7.2.1

Publications: [14]

Abstract: In this work, we present a numerical analysis of a method which combines a deterministic and a probabilistic approaches to quantify the migration of a contaminant, under the presence of uncertainty on the permeability of the porous medium. More precisely, we consider the flow equation in a random porous medium coupled with the advection-diffusion equation. Quantities of interest are the mean spread and the mean dispersion of the solute. The means are approximated by a quadrature rule, based on a sparse grid defined by a truncated Karhunen-Loève expansion and a stochastic collocation method. For each grid point, the flow model is solved with a mixed finite element method in the physical space and the advection-diffusion equation is solved with a probabilistic Lagrangian method. The spread and the dispersion are expressed as functions of a stochastic process. A priori error estimates are established on the mean of the spread and the dispersion.

### 6.4.5. An adaptive sparse grid method for elliptic PDEs with stochastic coefficients

**Participants:** Jocelyne Erhel, Mestapha Oumouni.

Grants: HYDRINV 7.4.5 , H2MN04 7.2.1

Publications: [31].

Abstract: The stochastic collocation method based on the anisotropic sparse grid has become a significant tool to solve partial differential equations with stochastic inputs. The aim is to seek a vector of weights and a convenient level of interpolation for the method. The classical approach uses an a posteriori approach on the solution, which causes an additional prohibitive cost.

In this work, we discuss an adaptive approach of this method to calculate the statistics of the solution. It is based on an adaptive approximation of the *inverse* diffusion parameter. We construct an efficient error indicator which is an upper bound of the error on the solution. In the case of unbounded variables, we use an appropriate error estimation to compute suitable weights for the method. Numerical examples are presented to confirm the efficiency of the approach, and to show that the cost is considerably reduced without loss of accuracy.

### 6.4.6. Numerical analysis of stochastic advection-diffusion equation via Karhunen-Loève expansion

**Participants:** Jocelyne Erhel, Mestapha Oumouni.

Grants: HYDRINV 7.4.5 , H2MN04 7.2.1

Publications: [32] , [25]

Abstract: In this work, we present a convergence analysis of a probabilistic approach to quantify the migration of a contaminant, under the presence of uncertainty on the permeability of the porous medium. More precisely, we consider the flow problem in a random porous medium coupled with the advection-diffusion equation and we are interested in the approximation of the mean spread and the dispersion of the solute. The conductivity field is represented by a Karhunen-Loève (K-L) decomposition of its logarithm. The flow model is solved using a mixed finite element method in the physical space. The advection-diffusion equation is computed thanks to a probabilistic Lagrangian method, where the concentration of the solute is the density function of a stochastic process. This process is solution of a stochastic differential equation (SDE), which is discretized using an Euler scheme. Then, the mean of the spread and dispersion are expressed as functions of the approximate stochastic process. A priori error estimates are established on the mean of the spread and of the dispersion. Numerical examples show the effectiveness of this approach.

### 6.4.7. About a generation of a log-normal correlated field

**Participants:** Jocelyne Erhel, Mestapha Oumouni.

Grants: HYDRINV 7.4.5 , H2MN04 7.2.1

Software: GENFIELD 5.2.3

Publications: in preparation

Abstract: Uncertainty quantification often requires the generation of large realizations of stationary Gaussian random field over a regular grid.

This paper compares and analyzes the commonalities between the methods and approaches for simulating stationary Gaussian random field. The continuous spectral method is the classical approach which discretizes its spectral density to construct an approximation of the field. When the spectral density and the covariance functions decrease rapidly to zero at infinity, we prove that the spectral method is computationally attractive.

We compare also the classical methods used to simulate the field defined by its covariance function, namely the Discrete Spectral method, the Circulant Embedding approach, and the Discrete Karhunen-Loève approximation. We have found that under some assumptions on the covariance all these latter methods give the same simulations of a stationary Gaussian field on a regular grid, which are very efficient with the Fast Fourier Transform algorithm.

### *6.4.8. A global model for reactive transport*

**Participants:** Édouard Canot, Jocelyne Erhel.

Grants: H2MN04 7.2.1 , MOMAS 7.2.5 , C2SEXA 7.2.3

Software: GRT3D 5.2.1

Thesis: [11]

Publications: [19], [16]

Abstract: In some scientific applications, such as groundwater studies, several processes are represented by coupled models. For example, numerical simulations are essential for studying the fate of contaminants in aquifers, for risk assessment and resources management. Chemical reactions must be coupled with advection and dispersion when modeling the contamination of aquifers. This coupled model combines partial differential equations with algebraic equations, in a so-called PDAE system, which is nonlinear. A classical approach is to follow a method of lines, where space is first discretized, leading to a semi-discrete differential algebraic system (DAE). Several methods have been designed for solving this system of PDAE.

In this study, we propose a global method which uses a DAE solver, where time is discretized by an implicit scheme. Then, each time step involves a nonlinear system of equations, solved by a modified Newton method. Thanks to the DAE solver, the time step is adaptively chosen in order to ensure accuracy and convergence. Moreover, the Jacobian in the nonlinear iterations is freezed as long as Newton converges fast enough, saving a lot of CPU time.

However, the size of the nonlinear system is quite large, because it involves both the differential and the algebraic variables. We show how to eliminate the differential variables, in order to reduce the size. This is equivalent to a so-called Direct Substitution Approach, but it keeps the nice features of DAE solvers.

Classicaly, the concentrations of chemical species are defined with their logarithms, assuming that they are strictly positive. This simplifies the computation of the mass action laws in the chemistry model and the computation of their derivatives. However, when a species does not exist, its concentration is replaced by a very small value and this may lead to an ill-conditioned Jacobian. We propose to use directly the concentrations, without logarithms, so that the Jacobian is then well-conditioned. Therefore, Newton method converges much faster without logarithms, allowing larger time steps and saving many computations.

We illustrate our method with two test cases, provided by the french agency for nuclear waste (ANDRA) and by the group MOMAS. We can compare our results with either analytical or other numerical solutions and show that our method is quite accurate. We also show that reducing the number of unknowns is very efficient and that dealing without logarithms reduces drastically the CPU time.

### *6.4.9. A chemistry model with precipitation-dissolution*

**Participant:** Jocelyne Erhel.

Grants: H2MN04 7.2.1 , MOMAS 7.2.5

Internship: Tangi Migot (Master M2, INSA and University, Rouen)

Publications: [36]

Abstract: In this study, we focus on precipitation and dissolution chemical reactions, because they induce numerical difficulties.

We consider a set of solute species and minerals, with precipitation occuring when a saturation threshold is reached. A challenge is to detect which minerals are dissolved and which minerals are precipitated. This depends on the total quantities of chemical species. We propose an analytical approach to build a phase diagram, which provides the interfaces between the different possible cases. We illustrate our method with three examples arising from brine media and acid mine drainage.

### *6.4.10. Coupled models for salted aquifers*

**Participants:** Édouard Canot, Jocelyne Erhel.

Grants: H2MN04 7.2.1 , MOMAS 7.2.5 , HYDRINV 7.4.5

Software: GEODENS and SELSAUM (from Tunis)

Internship: Marwen ben Refifa (Ph-D, ENIT, Tunis)

Publications: in preparation

Abstract: We study gravity driven problems in salted aquifers, when many species are present together with high concentrations. In this framework, we couple flow, transport and chemistry by using a fixed point approach. We interfaced two codes developed in Tunis: GEODENS for density driven flow and transport, and SELSAUM for geochemistry. This latter provides also the density of salted water.

## 6.5. Models and simulations for flow in porous fractured media

### *6.5.1. A Graphical User Interface for simulating flow and transport in fractured-porous media*

**Participants:** Jean-Raynald de Dreuzy, Jocelyne Erhel, Géraldine Pichot.

Grants: H2MNO4 7.2.1

Platform: H2OLab

Publications: [21]

Abstract: The platform H2OLab can be used with a Graphical User Interface, called H2OGuilde, which is developed using a Qt framework. Launchers correspond to a main program and to a hydrogeological application. These launchers call modules or libraries, implementing discretization schemes, solving algorithms, parallel communications, etc. The interface is generic for all the launchers. It is composed of three main tabs corresponding to the three steps of a simulation: entering input data, running computations, analyzing output data. Input parameters are classified in several categories, corresponding to the physical model and the numerical algorithms chosen. Output parameters are of three types, scalar, vector and matrix. Currently, visualization is done outside of the interface.

### *6.5.2. Meshing Strategies and the Impact of Finite Element Quality on the Velocity Field in Fractured Media*

**Participants:** Jean-Raynald de Dreuzy, Jocelyne Erhel, Géraldine Pichot.

Grants: FRACINI 7.1.1

Platform: H2OLab

Publications: [20]

Abstract: For solving flow within a network of fractures, Mixed Hybrid Finite Element (MHFE) method is a method of choice as it yields a symmetric positive definite linear system. However, a drawback is its sensitivity to bad aspect ratio elements. For poor-quality triangles, elementary matrices are ill-conditioned and inconsistent velocity vectors are obtained by inverting these local matrices. In our presentation, we will present different strategies for a better reconstruction of the velocity field.

## SERPICO Project-Team

# 6. New Results

## 6.1. Patch-based statistical denoising methods for electron and light microscopy

**Participants:** Charles Kervrann, Frédéric Lavancier.

Inspired form the non-local means [33], we developed a stochastic NL-means-based denoising algorithm for generalized non-parametric noise models [21], [9]. First, we provided a statistical interpretation to current patch-based neighborhood filters and justify the Bayesian inference that needs to explicitly account for discrepancies between the model and the data. Furthermore, we investigated the Approximate Bayesian Computation (ABC) rejection method [30], [46] combined with density learning techniques for handling situations where the posterior is intractable or too prohibitive to calculate. This is particularly relevant for images contaminated by heterogeneous sources of noise. A major difference with previous methods is that we directly handle the structure of the noise, without precise parametric modeling of the noise. We demonstrated the flexibility of our stochastic Gamma non-local means (SGNL-means) by showing how it can be adapted to tackle noise in frequency domain fluorescence lifetime imaging microscopy (FD-FLIM) and cryo-electron tomography (see Fig. 3 ).
Moreover, we also proposed a general statistical aggregation method which combines image patches denoised with several commonly-used algorithms [20]. We showed that weakly denoised versions of the input image obtained with standard methods, can serve to compute an efficient patch-based aggregated estimator. In our approach, we evaluate the Stein's Unbiased Risk Estimator (SURE) of each denoised candidate image patch and use this information to compute the exponential weighted aggregation (EWA) estimator. The aggregation method is flexible enough to combine any standard denoising algorithm and has an interpretation with Gibbs distribution. The denoising algorithm (PEWA) is based on an MCMC sampling and is able to produce results that are comparable to the state of the art ([4], [38]). In this range of work, we have also introduced in [28] a general method to combine estimators in order to produce a better estimate. From a theoretical point of view, we proved that this method is optimal in some sense. It is illustrated on standard statistical problems in parametric and semi-parametric models where the averaging estimator outperforms the initial estimators in most cases. As part of an on-going work, we are applying this method to improve patch-based image denoising algorithms.

**References:** [9] [21] [20] [28]

**Collaborators:** Philippe Roudot (UT Southwestern Medical Center, Dallas (TX))
Francois Waharte (UMR 144 CNRS-Institut Curie, STED team and PICT-IBiSA)
Paul Rochet (Laboratoire de Mathématiques Jean Leray (LMJL), university of Nantes)

## 6.2. Design of deconvolution algorithms for low exposure fluorescence microscopy images

**Participants:** Deepak George Skariah, Charles Kervrann.

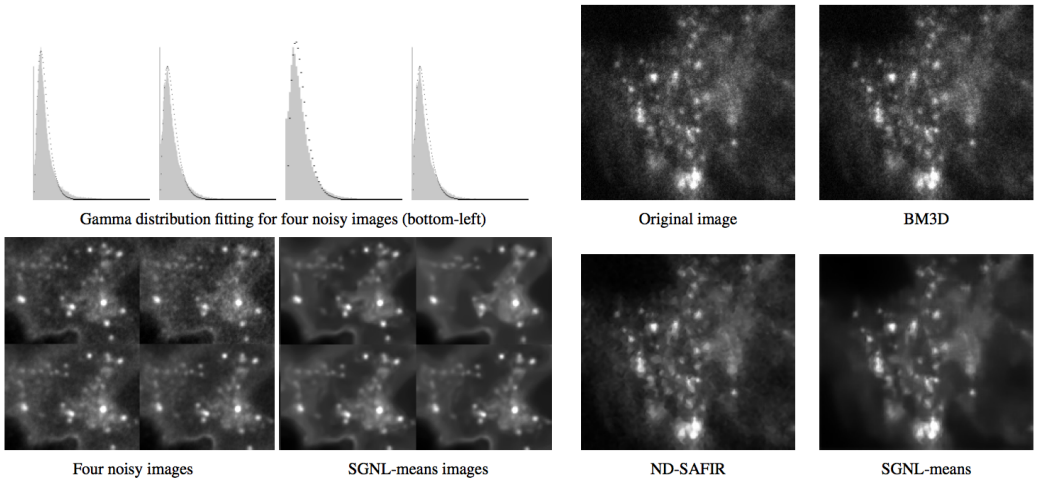


*Figure 3. Experiments in FD-FLIM (confocal spinning-disk microscopy, UMR 144 CNRS-Institut Curie, PICT-IBiSA). Left: FNAR1 tagged with Green Fluorescence Protein (GFP) observed in a epithelial cell with mCHerry-tagged Tyk2 ; Gamma distribution fitting and SGNL-means denoising on four successive images with temporally varying signal-to-noise ratios. Right: comparison of denoised images with methods [4] [38].*

Fluorescence imaging is popular in cell biology research due to its high contrast imaging capability. In microscopy imaging under low exposure conditions, the image quality is limited by out-of-focus blur and high noise. As a result a preprocessing stage known as deconvolution is needed to estimate a good quality version of the observed image. We proposed to design an efficient deconvolution algorithm for fluorescence microscopy under low exposure conditions by using the Poisson noise model. The result of deconvolution depends heavily on the choice of the regularization term. The regularization functional should be designed to remove noise while retaining the image structure. The choice of Poisson noise model and new regularization functional demands the design of a new and efficient optimization algorithm. We proposed to use a complex non quadratic regularization functional along with Poisson noise assumption for the first time. The use of non quadratic regularization makes the resulting optimization problem a complex one. This demanded the development of a problem-specific optimization algorithm which is fast as well as robust enough to minimize a non quadratic cost function. The use of non quadratic regularization together with Poisson noise model ensures that finer details of underlying structures are well restored in the presence of large amount of noise.

**Collaborator:** Muthuvel Arigovindan (Imaging Systems Lab, Department of Electrical Engineering, Indian Institute of Science, Bangalore, India).

## 6.3. Background estimation and vesicle segmentation in live cell imaging

**Participants:** Thierry Pécot, Patrick Bouthemy, Charles Kervrann.

In live cell fluorescence microscopy images, the moving tagged structures of interest, such as vesicles, often appear as bright spots with intensity that varies along time over a time-varying and cluttered background. Localization and morphology assessment of these small objects over time is then crucial to provide valuable information for quantitative traffic analysis. In this study, we have focused on the Rab6 protein as a typical intracellular membrane-associated protein. Rab6 is known to promote vesicle trafficking from Golgi to Endoplasmic Reticulum or to plasma membrane. In our study, micro-fabricated patterns have been used to enforce cells to have circular or crossbow normalized shape. Micro-patterns impose constraints on the

cytoskeleton and the location of organelles (e.g. Golgi apparatus) is thus better controlled. These micro-patterns also influence the spatial distribution of Rab6 transport carriers. However, the direct influence of the micro-patterns on the spatial dissemination of these trafficking vesicles has so far not been completely characterized. In this work, we have considered a statistical Bayesian approach in the framework of conditional random fields (CRF) for background estimation and vesicle segmentation [13]. Within this approach, we have designed a robust detection measure for fluorescence microscopy based on the distribution of neighbor patch similarity. We formulate the vesicle segmentation and background estimation as a global energy minimization problem. An iterative scheme to jointly segment vesicles and background is proposed for 2D-3D fluorescence image sequences. We have conducted a quantitative comparison with state-of-the-art methods on a large set of synthetic image sequences with a cluttered time-varying background and achieved a quantitative validation of the vesicle segmentation method on 2D and 3D micro-patterned cells expressing GFP-Rab6.

**Reference:** [13]

**Collaborators:** Jean Salamero (UMR 144 CNRS-Institut Curie, STED team and PICT-IBiSA)
Jérôme Boulanger (UMR 144 CNRS-Institut Curie, STED team)

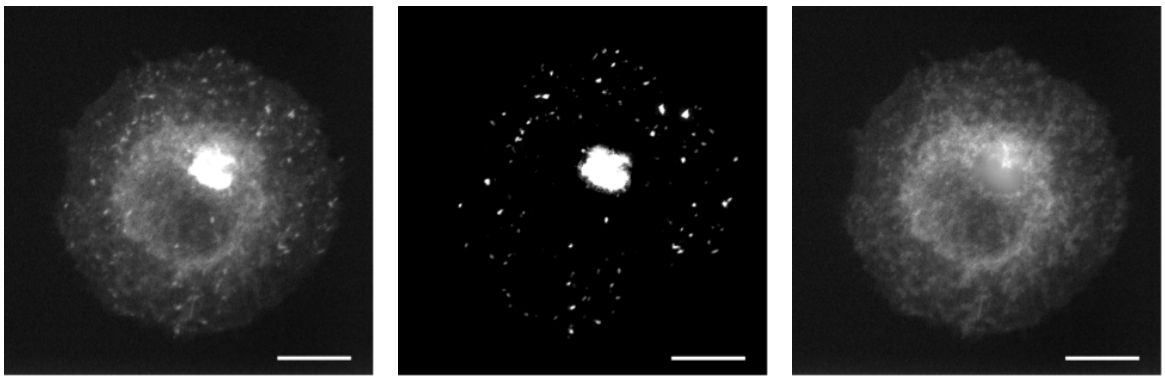

*Figure 4. Left: Fuorescence confocal spinning-disk microscopy image depicting GFP-Rab6 proteins (UMR 144 CNRS-Institut Curie, PICT-IBiSA). Middle: estimated vesicular component. Right: estimated background.*

## 6.4. A quantitative approach for space-time membrane trafficking orientation

**Participants:** Thierry Pécot, Patrick Bouthemy, Charles Kervrann.

Rab6 proteins are trafficking from the Golgi apparatus at the cell center to Endoplasmic Reticulum or to plasma membrane located at the periphery of the cell. The cell shape influences Rab6 trafficking but no study has ever quantified the effect of the cell shape on the trafficking orientation. In this study [25], we compare Rab6 trafficking orientation constrained by two different micropatterns [56] (circular and crossbow-shaped cells) from fluorescence video-microscopy. Object/background separation [13] is first applied to 3D+T image sequences to extract Rab6 spatio-temporal coordinates. The bandwidth of the von Mises kernel is automatically estimated using the rule of thumb and leads to two different densities for the two different micropatterns. We propose to quantitatively compare these densities by computing the Wilcoxon rank sum paired test between inter- and intra-micropattern distances. We considered the circular earth mover's distance (also known as the Wasserstein metric) to compare traffic densities. Our quantitative study on micro-patterned cells concludes that the Rab6 transport carriers destinations concentrate at the three corner points of the crossbow-shaped cells corresponding to the main adhesion sites, while the vesicle destination distribution is somewhat uniform for circular-shaped cells.

**Reference:** [25]

**Collaborators:** Jean Salamero (UMR 144 CNRS-Institut Curie, STED team and PICT-IBiSA)
Jérôme Boulanger (UMR 144 CNRS-Institut Curie, STED team)

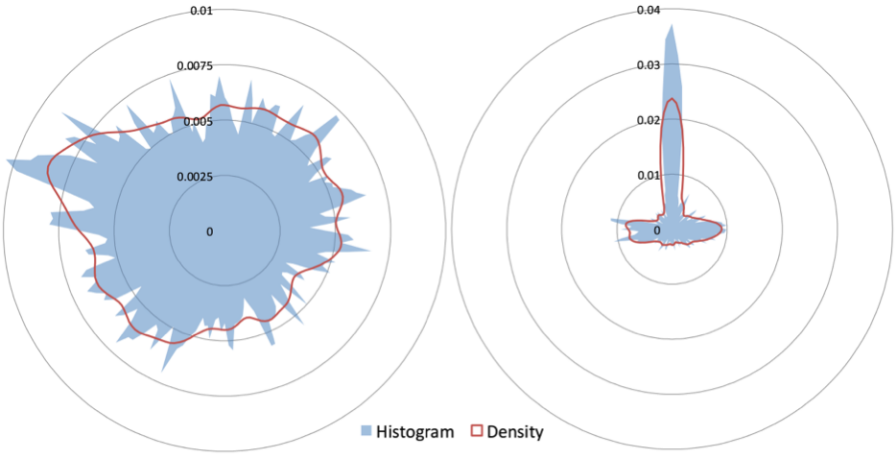


*Figure 5. Distribution of traffic orientation for circle-shaped cells (left) and crossbow-shaped cells (right).*

## 6.5. Vesicle segmentation method with automatic scale selection in TIRF microscopy

**Participants:** Antoine Basset, Charles Kervrann, Patrick Bouthemy.

Accurately detecting subcellular particles in fluorescence microscopy is of primary interest for further quantitative analyses such as counting, tracking or classification. Our primary goal was to segment vesicles in fluorescence microscopy images. In [15] we proposed a first spot detection method with automatic scale selection. We have now dramatically improved the precision of the scale selection step, yielding to a more reliable detection of the spots [23]. The method relies on a Laplacian of Gaussian (LoG) filter to first enhance the spots while reducing noise. To obtain good detection results, the scale of the Gaussian filter must be precisely set, according to the spots size [23]. In order to cope with very small spots, we rely on the discrete analog of the Gaussian filter [45], instead of the previously used sampled Gaussian filter. With this filter, we can find the optimal Gaussian scale with an arbitrary precision by minimizing a statistical criterion. We have introduced two criteria for this purpose and compared them. Once the optimal scale is selected, we threshold the lowest values of the LoG-filtered image, which correspond to spots. To cope with inhomogeneous background, thresholding must be adapted to local statistics so that a single probability of false alarm (PFA) setting can be defined for the whole image or even the collection of images to be processed. In short, we automatically infer from image data the optimal parameters usually left to the user guidance in other methods, that is, spot scale and detection threshold. We have carried out an extensive comparative evaluation, which demonstrates that our new scale selection approach improves detection performances, and that our spot detection method outperforms state-of-the-art detectors [23].

**References:** [15] [23]

**Collaborators:** Jean Salamero (UMR 144 CNRS-Institut Curie, STED team and PICT-IBiSA)
Jérôme Boulanger (UMR 144 CNRS-Institut Curie, STED team)

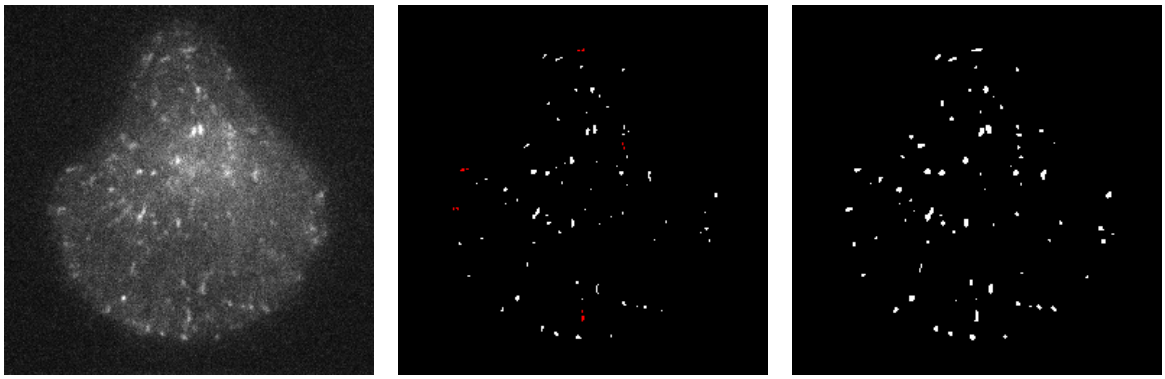

*Figure 6. Comparison of segmentation results on a real image presenting elongated spots. Left: Input TIRFM images (Rab11-mCherry) (UMR 144 CNRS-Institut Curie, PICT-IBiSA). Middle: Segmentation results with state-of-the-art detector MS-VST [60]. Some elongated spots of (left) are split (red) by MS-VST due to a too small filter scale. Right: Segmentation results with our new detection method. Elongated objects are well recovered thanks to the precise scale selection.*

## 6.6. Analysis of the repartition of moving vesicles by spatio-temporal point process models

**Participants:** Frédéric Lavancier, Thierry Pécot, Charles Kervrann.

Characterizing the spatial repartition of interacting moving proteins is a fundamental step for co-localization and co-expression. Based on the segmentation algorithm [15], [23], this challenge amounts to characterizing the repartition or spatial distribution of spots (see Fig. 6 ). This is part of the more general statistical analysis of random geometrical objects, and in particular of random points. Gibbs models form a large class of point process models, that can be used to characterize either complete randomness or attraction or repulsion between points depending on the Gibbs potential at hand.

First in [27], we focused on infinite range potentials that include the most famous interaction potential arising from statistical physics, namely the Lennard Jones potential. To fit this kind of models to a dataset, the standard inference methods are not applicable. We introduced in [27] a modification of the pseudolikelihood method, with a specific border correction, and we prove that this provides consistent and asymptotically normal estimators. Second, in [26], we studied an alternative class of models, the determinantal point processes (DPP). They are designed to model repulsion between points and are thus adapted to regular point patterns. These models are becoming very popular in the spatial statistics community due to many appealing properties. We quantified the possible repulsiveness that a DPP can model [26]. In particular, we determined the most repulsive stationary DPP. We finally introduced new parametric families of DPPs that cover a large range of DPPs, from the homogeneous Poisson process (for no interaction) to the most repulsive DPP.

An application of these models to the problem of co-localization between proteins is part of an on-going project. In each protein, the set of vesicles is modeled by a union of random balls, possibly overlapping, and a Gibbs interaction is introduced to take into account the possible interaction in the location of vesicles between two proteins. Our first concern is to test whether the two proteins actually interact, i.e. co-localization occurs, or in other words whether the Gibbs interaction is empty or not. If there is co-localization, the further step is to characterize it through the estimation of the strength of the Gibbs interaction.

**References:** [26] [27]

**Collaborators:** Christophe Ange Napoléon Biscio (LMJL, University of Nantes)

Jean-François Coeurjolly (Laboratoire Jean Kuntzmann, Grenoble Alpes University)

## 6.7. Detection and estimation of membrane diffusion during exocytosis in TIRF microscopy

**Participants:** Antoine Basset, Charles Kervrann, Patrick Bouthemy.

Assessing the dynamics of plasma membrane diffusion processes in live cell fluorescence microscopy is of paramount interest to understand cell mechanisms. We investigated methods to detect vesicle fusion events, and estimate the associated diffusion coefficients in TIRFM image sequences [16]. In contrast to classical approaches, a diffusion coefficient is locally estimated for each detected fusing vesicle. We first detect the membrane fusion events and then select the diffusion configurations among them with a correlation test. To estimate the diffusion coefficient, a geometric model is fitted to the detected spot directly in the 2D+T subvolume. This recent estimation approach produced more satisfying results when compared to [16]. Diffusion events are reliably recognized, and the diffusion coefficient is accurately estimated for each diffusion event. This work will be integrated in a broader study, spanning from transport phase to membrane fusion, and non-diffusion events will be analyzed.

**Reference:** [16]

**Collaborators:** Jean Salamero (UMR 144 CNRS-Institut Curie, STED team and PICT-IBiSA)
Jérôme Boulanger (UMR 144 CNRS-Institut Curie, STED team)

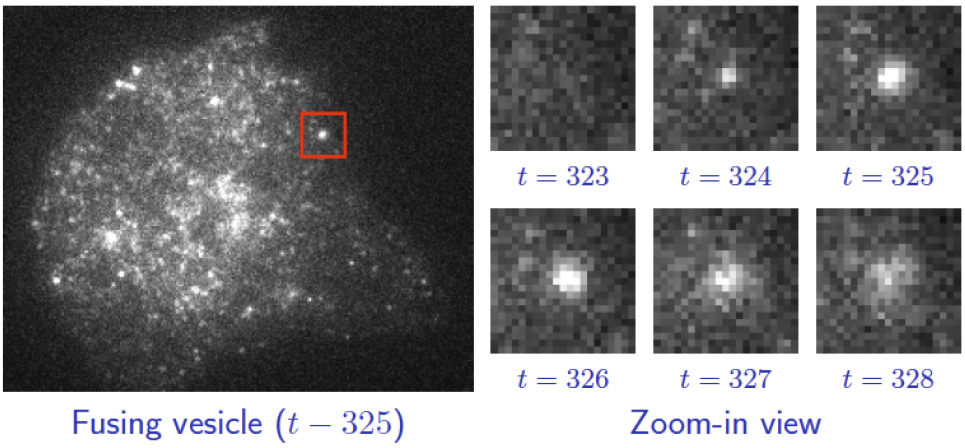


*Figure 7. Left: Fusing vesicle (frame in red) in a TIRFM (UMR 144 CNRS-Institut Curie, PICT-IBiSA) sequence (frame 325, 50ms/frame). Right: Zoom-in view of the temporal evolution of the fusing vesicle.*

## 6.8. Estimation of the flow of particles without tracking in fluorescence video-microscopy

**Participants:** Thierry Pécot, Patrick Bouthemy, Charles Kervrann.

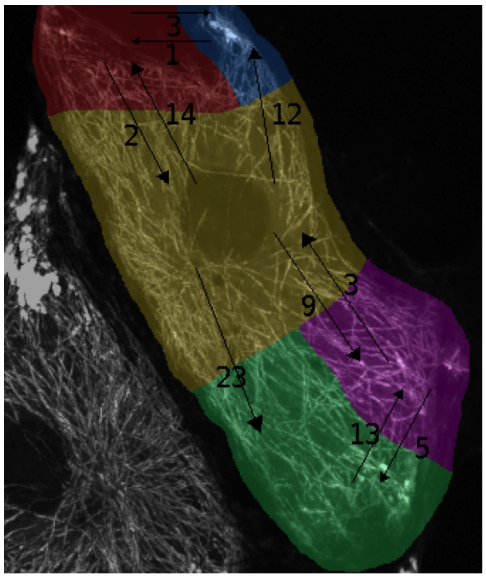


*Figure 8. Vesicle flows estimated when considering a simple partition of 5 regions for an image sequence acquired in TIRF microscopy and showing the protein Clip170 (UMR 144 CNRS-Institut Curie, PICT-IBiSA).*

Automatic analysis of the dynamic content in fluorescence video-microscopy is crucial for understanding molecular mechanisms involved in cell functions. We have proposed an original approach for analyzing particle trafficking in these sequences. Instead of individually tracking every particle, we only locally count particles crossing boarders between regions over time and minimize a global energy function. Three methods to determine the particle flow have been considered. We have conducted comparative experiments on synthetic and real fluorescence image sequences. We have shown that adding a sparsity constraint on the number of detected events allows us to reduce the number of false alarms. Compared to usual tracking methods, our approach is simpler and the results are very stable. This estimation method needs the adjustment of only two parameters. (see Fig. 8 ).

**Reference:** [22]

**Collaborators:** Jean Salamero (UMR 144 CNRS-Institut Curie, STED team and PICT-IBiSA)
Jérôme Boulanger (UMR 144 CNRS-Institut Curie, STED team)

## 6.9. Detection and tracking of astral microtubules at the cell cortex

**Participants:** Thierry Pécot, Charles Kervrann, Geoffrey Dieffenbach.

In this study, we are interested in the influence of the mechanical properties of astral microtubules in the centering mechanisms of the mitotic spindle, giving it a robust positioning. In their previous studies, the CeDRE group (IGDR Rennes) identified two subpopulations of astral microtubules that either push or pull the cell cortex. To better understand these mechanisms, image sequences are acquired at the cortex level where extremities of astral microtubules come to exert forces. In order to characterize the two subpopulations of astral microtubules during the mitosis in the unicellular embryos of C. Elegans, life span, that is the period during which the microtubule is touching the cell cortex, for every single microtubule has to be measured. A short life span corresponds to a pulling force while a longer life span corresponds to a pushing force. Detecting and tracking microtubules at the cell cortex has to be done to collect these measures. As the signal-to-noise ratio is low, a denoising step is needed to detect the microtubule extremities. Several detection methods were

tested but we need to further investigate this step to find the most suited methods for this particular application. Finally, the U-track algorithm [42] is applied to track the microtubules extremities to measure their life span.

**Collaborators:** Jacques Pécréaux (CeDRE group, IGDR Rennes, CNRS UMR 6290)
Hélène Bouvrais (CeDRE group, IGDR Rennes, CNRS UMR 6290)

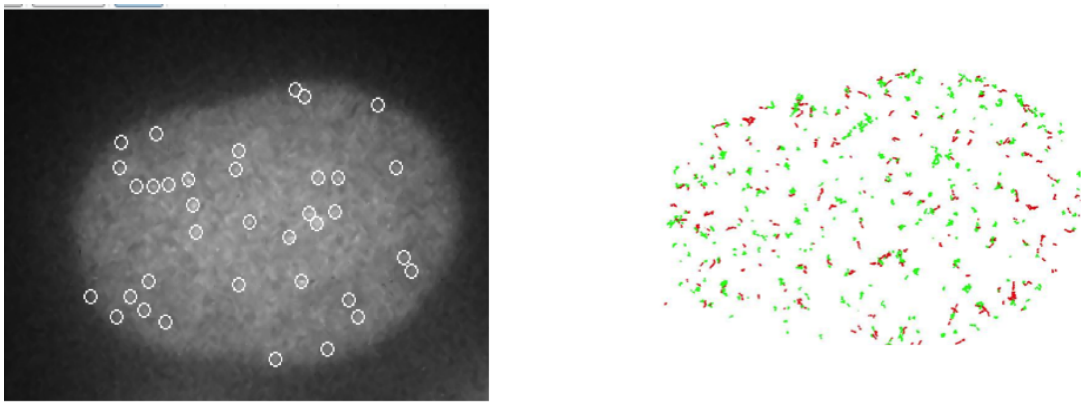

*Figure 9.*

*Microtubule extremities detection and tracking in fluorescence microscopy (embryo of C. Elegans, IGDR - Institute of Genetics and Developmental biology of Rennes, CNRS UMR 6290).*

## 6.10. Spot localization and segmentation for Tissue MicroArray (TMA) de-arrying

**Participants:** Hoai Nam Nguyen, Charles Kervrann.

Tissue core de-arraying is one of the most important steps in tissue microarray (TMA) image analysis. A very first task of TMA (Tissue MicroArray) image analysis is to accurately localize spots (separate tissue core) representing arrays of $512 \times 512$ pixels each, in very large images of several thousands of pixels. However, few solutions and frameworks are available and none of them covers images provided by fluorescent scanners. We developed a robust TMA de-arraying method adapted for digital images from classical optical and new fluorescent devices. The proposed algorithm is composed of three modules: i) detection, ii) segmentation, and iii) array indexing. The detection of TMA cores is performed by local adaptive thresholding of isotropic wavelet transform coefficients. We demonstrated how a wavelet decomposition at any desired scale can be performed faster than usual techniques by exploiting explicit formula of the analysis wavelet. Our core detection strategy enables to deal with images having significant noise level, inhomogeneous background, and high dynamic range such as fluorescence images, without any assumption on image noise and intensity value range. The detected cores are furhter segmented by using parametric ellipse model to improve detection accuracy. Combining these two modules, we can handle complex background and artifacts, particularly in fluorescence imaging, and thus reduce false detections. After the segmentation step, the position of detected cores is determined by the centroid of relevant segments. Finally, to compute array indices of cores, we estimate the deformation of a theoretical grid under a thin-plate model by using an iterative scheme. After each iteration, the initial regular grid is progressively transformed for fitting computed core positions. Our main contribution is the reformulation of the array indexing problem as an estimation of the deformation function, which is solved with a iterative algorithm. Moreover, when design layout of TMA slide is known, our estimator of deformation yields quantitative information about grid deformation such as average translation, rotation angle, shearing coefficients, bending energies along axis, etc. They can be used as quality indicators of the manufactured TMA slide.

**Collaborator:** Vincent Paveau (Innopys company)

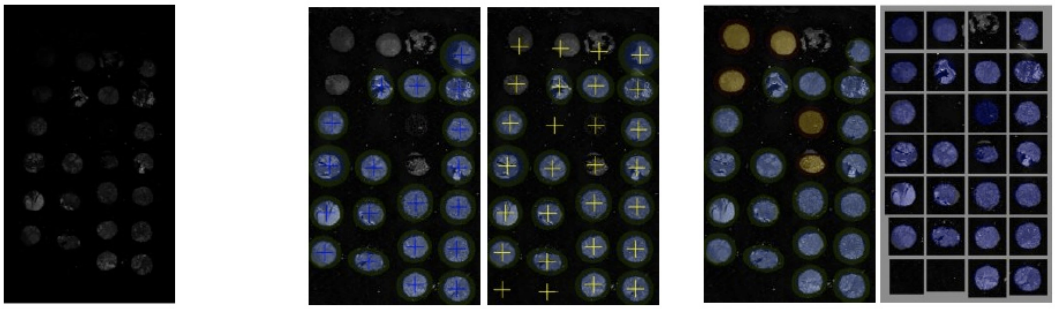

*Figure 10. Array indexing TMA (Innopsys company). From left to right : input TMA image, segmented core positions marked by blue crosses, estimated positions of deformed grid marked by yellow crosses, retrieved missed cores after detection/segmentation steps (orange areas), and array representation of TMA (retrieved cores are colored).*

## 6.11. Adaptive global and local motion estimation

**Participants:** Noémie Debroux, Charles Kervrann.

The design of data costs is one of the main research issue for variational optical flow estimation. The aim is to improve discriminative power by integrating appropriate neighborhood information, while preserving computational efficiency. Most previous works define features on patches with predefined sizes and shapes, or filter pixelwise costs with fixed filtering parameters. We proposed a novel approach estimating spatially varying parameters of filters used to define the data term [8]. More specifically, our model considers Gaussian filtering of the pixelwise brightness constancy equation and imposes smoothness constraints on motion and convolution filter size (bandwith). The energy encoding these assumptions is alternatively minimized over flow field and the spatially varying bandwidth in a variational framework. Experimental results on the Middlebury database demonstrated clear improvements yielded by our method over the spatially constant case of [32] (see Fig. 11 ).

**Collaborator:** Denis Fortun (UMR 144 CNRS-Institut Curie, STED team, Paris)
(EPFL, Lausanne, Switzerland)

## 6.12. Crowd motion classification

**Participants:** Antoine Basset, Charles Kervrann, Patrick Bouthemy.

Assessing crowd behaviors from videos is a difficult task while of interest in many applications. We have defined a novel approach which identifies from two successive frames only, crowd behaviors expressed by simple image motion patterns. It relies on the estimation of a collection of sub-affine motion models in the image, a local motion classification based on a penalized likelihood criterion, and a regularization stage involving inhibition and reinforcement factors [17]. The apparent motion in the image of a group of people is assumed to be locally represented by one of the three following motion types: translation, scaling or rotation. The three motion models are computed in a collection of predefined windows with the robust estimation method [48]. At every point, the right motion model is selected owing to the corrected (for small sample size) Akaike information criterion (AICc). To classify the local motion type, the three motion models are further subdivided into a total of eight crowd motion classes. Indeed, scaling refers either to gathering (Convergence) or dispersing people (Divergence). Rotation can be either Clockwise or Counterclockwise.

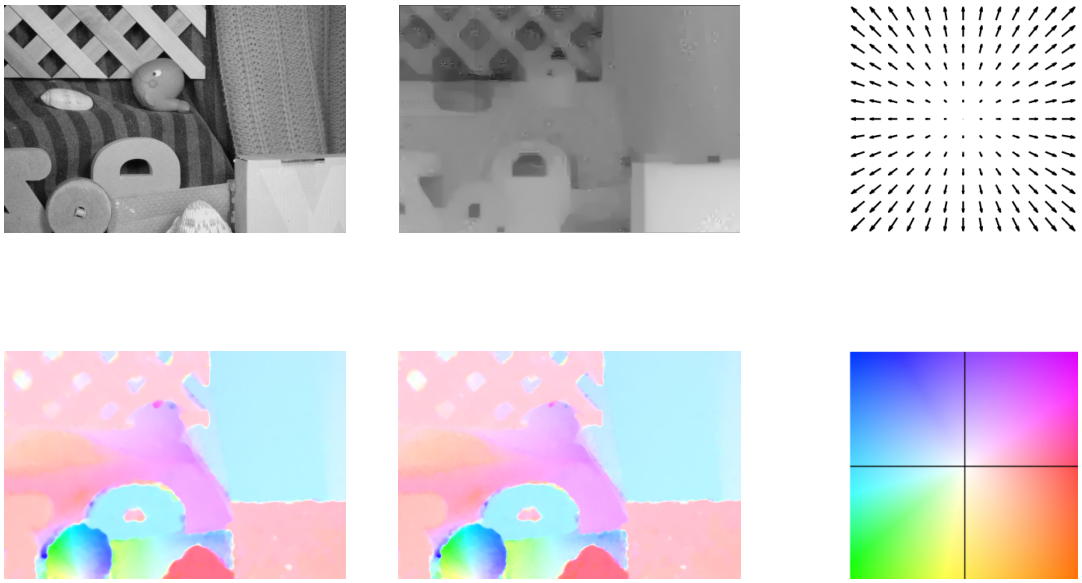

*Figure 11. Comparison on a sequence of the Middlebury benchmark. Top from left to righ: input image and spatially filter bandwidth estimation. Bottom from left to right: velocity field computed by [32] (endpoint error = 0.143) and by our method (endpoint error = 0.126).*

Since our classification scheme is view-based, four image-related translation directions are distinguished: North, West, South, East. Then, to get the final crowd classification, a regularization step is performed, based on a decision tree and involving inhibition for opposed classes such as convergence and divergence. We have also developed an original and simple method for recovering the dominant paths followed by people in the observed scene. It involves the introduction of local paths determined from the space-time average of the parametric motion subfields selected in image blocks. Starting from one given block in the image, we straightforwardly reconstruct a global path by concatenating the local paths from block to block. Experiments on synthetic and real scenes have demonstrated the performance of our method, both for motion classification and principal paths recovery.

**Reference:** [17]

## 6.13. Anomaly detection using block-based histograms of crowd motion patterns

**Participants:** Juan Perez Rua, Antoine Basset, Patrick Bouthemy.

We have developed a new and generic method to detect and localize abnormal events in videos of crowd scenes. The algorithm consists first in determining the flow vector and crowd motion class for every moving pixel from a set of affine motion models estimated on a collection of windows. Then, the observed scene is subdivided in blocks to compute crowd motion class histograms weighted by the motion vector magnitudes. A very simple training step enables to get the reference histograms per block accounting for the normal behaviours. For each block, we can automatically set by means of statistical arguments the threshold on the distance between the histogram in the current image and the reference histogram that decides the presence of an abnormal event in that block. Results of extensive experimentation on different types of anomaly datasets show that our method is competitive with respect to methods relying on far more elaborated models on both appearance and motion and thus involving a significant learning stage. It outperforms any other existing purely motion-based anomaly localization method.

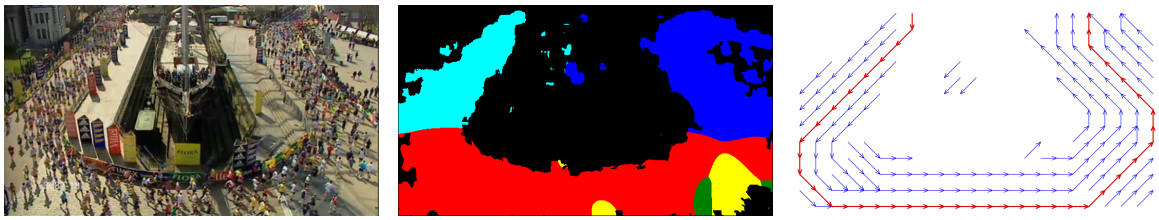

*Figure 12. Overview of the method applied to a sequence where runners follow a 'U' from the upper left corner to the upper right corner. Left: First frame of the sequence. Middle: Classification results (cyan=translation toward South, red=counterclockwise rotation, yellow=translation toward East, green=convergence, blue=translation toward North). Right: Recovery of the longest path in the scene (red).*

**SIROCCO Project-Team**

# 6. New Results

## 6.1. Analysis and modeling for compact representation and navigation

3D modelling, multi-view plus depth videos, Layered depth images (LDI), 2D and 3D meshes, epitomes, image-based rendering, inpainting, view synthesis

### *6.1.1. Salient object detection*

**Participants:** Olivier Le Meur, Zhi Liu.

Salient object detection consists in extracting in an automatic manner the most interesting object in an image or video sequence. From an input image, an object, with well-defined boundaries, is detected based on its saliency. This subject knows an renewed interest these last years. A number of datasets serving as ground truth has been released and can be used to benchmark methods.

In 2013, we proposed a new method for detecting salient objects in still color images. In 2014, this method has been extended to video sequences [21]. Based on the superpixel representation of video frames, motion histograms and color histograms are computed at local and global levels. From these histograms, a superpixel-level temporal saliency measure as well as a spatial saliency measure are obtained. Finally, a pixel-level saliency derivation method is proposed to generate pixel-level temporal saliency map and spatial saliency map. An adaptive fusion method allows to integrate them into an unique spatiotemporal saliency map. Experimental results on two public datasets demonstrate that the proposed model outperforms state-of-the-art spatiotemporal saliency model in terms of both saliency detection and human fixation prediction.

### *6.1.2. Saliency aggregation*

**Participants:** Olivier Le Meur, Zhi Liu.

In this study [32], we investigate whether the aggregation of saliency maps allows to outperform the best saliency models. Today there exist a number of saliency models for predicting the most visually salient locations within a scene. Although all existing models follow the same objective, they provide results which could be, to some extent, different. The discrepancies are related to the quality of the prediction but also to the saliency map representation. Indeed some models output very focused saliency maps whereas the distribution of saliency values is much more uniform in other models. Others tend to emphasize more on the image edges, the color or luminance contrast. This saliency map manifold contains a rich resource that should be used and from which new saliency maps could be inferred. Combining saliency maps generated using different models might enhance the prediction quality and the robustness of the prediction. Our goal is then to take saliency maps from this manifold and to produce the final saliency map.

This study discussed various aggregation methods; six unsupervised and four supervised learning methods are tested on two existing eye fixation datasets. Results show that a simple average of the TOP 2 saliency maps significantly outperforms the best saliency models. Considering more saliency models tends to decrease the performance, even when robust aggregation methods are used. Concerning the supervised learning methods, we provide evidence that it is possible to further increase the performance, under the condition that an image similar to the input image can be found in the training dataset. Our results might have an impact for critical applications which require robust and relevant saliency maps.

### *6.1.3. Models for 3D video quality assessment*

**Participants:** Darya Khaustova, Olivier Le Meur.

This work is carried out in collaboration with Orange labs. The goal is to design objective metrics for quality assessment of 3D video content, by establishing links between human visual perception (visual comfort) and video parameters such as quality and depth quantity, and between visual comfort and visual attention. In 2013 we investigated the differences in 2D visual attention in comparison with 3D visual attention [31]. In 2014, we have focused on the design of an objective stereoscopic quality metric. In stereoscopic video quality, the assessment of spatial and temporal distortions by conventional quality metrics became incomplete because of the added depth dimension. Improperly captured or rendered, depth information can induce visual discomfort, impacting the overall video 3D QoE quality independently of image quality. The model is based on perceptual thresholds, namely visual annoyance, and acceptability. The visual annoyance threshold defines the boundary between annoying and not annoying sensation: 50% of subjects consider a stimulus annoying and 50% as not annoying. Acceptability determines the viewer's expectation level for the perceived video quality in a certain context and situation (inspired by the acceptability for the customer defined as an adequate service.

In order to compute the quality score, the proposed metric requires in input the distortion level of a technical and particular parameter, annoyance threshold and acceptability threshold of the targeted parameter. The performance of proposed objective mode is evaluated by considering five view asymmetries with five degradation levels. Generated contents were assessed by 30 subjects for each asymmetry (focal length mismatch, vertical shift, and rotation, green and white level reduction). The results of the subjective test have demonstrated that it is possible to classify detected problem to one of the objective categories using corresponding acceptability and visual annoyance thresholds.

### *6.1.4. Epitome-based video representation*

**Participants:** Martin Alain, Christine Guillemot.

In 2014, we have developed fast methods for constructing epitomes from images. An epitome is a factorized texture representation of the input image, and its construction exploits self-similarities within the image. Known construction methods are memory and time consuming. The proposed methods, using dedicated list construction on one hand and clustering techniques on the other hand, aim at reducing the complexity of the search for self-similarities. Experiments show that interesting complexity results can be obtained without degrading the epitome quality for both proposed methods. By limiting the number of exhaustive searches we limit the memory occupation and the processing time, while keeping a good epitome quality (down to 18.08 % of the original memory occupation and 41.39 % of the original processing time) [25]. As an example, images reconstructed using the different techniques are visible in Fig. 1 . The epitome construction method is currently being extended from still images to groups of images in video sequences. Denoising and super-resolution algorithms based on the constructed epitomes are also under study.

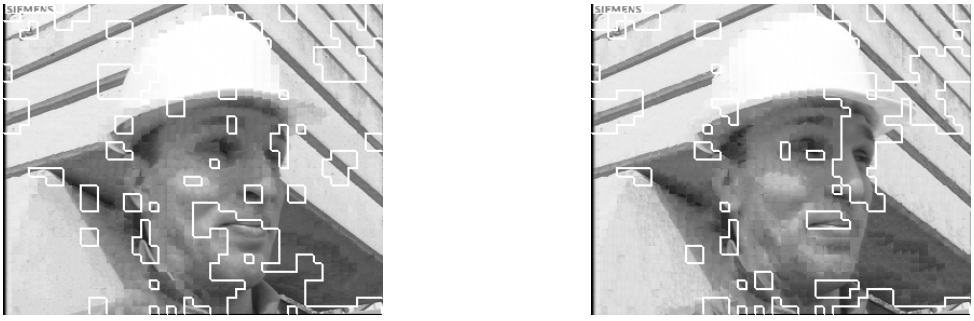

*Figure 1. Reconstructed images using the list-based (left) and clustering-based methods. Epitome patches are highlighted in white.*

### *6.1.5. Light field tomographic reconstruction from a fixed camera focal stack*

**Participants:** Christine Guillemot, Elif Vural.

Thanks to the internship of Antoine Mousnier (student at Ecole Centrale Lyon), we have developed a novel approach to partially reconstruct high-resolution 4D light fields from a stack of differently focused photographs taken with a fixed camera. First, a focus map is calculated from this stack using a simple approach combining gradient detection and region expansion with graph cut. Then, this focus map is converted into a depth map thanks to the calibration of the camera. We proceed after this with the tomographic reconstruction of the epipolar images by back-projecting the focused regions of the scene only. We call it masked back-projection. The angles of back-projection are calculated from the depth map. Thanks to the high angular resolution we achieve, we are able to render puzzling perspective shifts although the original photographs were taken from a single fixed camera at a fixed position and render images with extended focus (see Fig. 2 ). To the best of our knowledge, our method is the first one to reconstruct a light field by using a focal stack captured with an ordinary camera at a fixed viewpoint.

*Figure 2. Three images of the focal stack (left); estimated depth map and image with extended focus (right). The focal stack images of the first and second rows have been captured with a Nikon 5200 camera.*

## 6.2. Rendering, inpainting and super-resolution

image-based rendering, inpainting, view synthesis, super-resolution

### *6.2.1. Video inpainting*

**Participants:** Mounira Ebdelli, Christine Guillemot, Olivier Le Meur.

Image (and video) inpainting refers to the process of restoring missing or damaged areas in an image (or a video). This field of research has been very active over the past years, boosted by numerous applications: restoring images from scratches or text overlays, loss concealment in a context of impaired image transmission, object removal in a context of editing, disocclusion in image-based rendering of viewpoints different from those captured by the cameras. Inpainting is an ill-posed inverse problem: given observations, or known samples in a spatial (or spatio-tempoal) neighborhood, the goal is to estimate unknown samples of the region to be filled in. Many methods already exist for image inpainting, either based on PDE (Partial Derivative Equation)-based diffusion schemes, either using sparse or low rank priors or following texture synthesis principles exploiting statistical or self-similarity priors.

In 2014, the problem of video inpainting has been further addressed with free-moving cameras. The algorithm developed first compensates the camera motion between the current frame and its neighboring frames in a sliding window, using a new region-based homography computation which better respects the geometry of the scene compared to state-of-the-art methods. The source frame is first segmented into regions in order to find homogeneous regions. Then, the homography for mapping each region into the target frame is estimated. The overlapping of all aligned regions forms the registration of the source frame into the target one. Once the neighboring frames have been aligned, they form a stack of images from which the best candidate pixels are searched in order to replace the missing ones. The best candidate pixel is found by minimizing a cost function which combines two energy terms. One energy term, called the data term, captures how stationary is the background information after registration, hence enforcing temporal coherency. The second term aims at favoring spatial consistency and preventing incoherent seams, by computing the energy of the difference between each candidate pixel and its 4-neighboring pixels in the missing region. The minimization of the energy term is performed globally using Markov Random Fields and graph cuts. A method of Poisson blending has been implemented in order to further enhance the visual quality of the inpainted videos. The proposed approach, although less complex than state-of-the-art methods, provides more natural results.

### *6.2.2. Image and video super-resolution in the example-based framework*

**Participants:** Marco Bevilacqua, Christine Guillemot, Aline Roumy.

Super-resolution (SR) refers to the problem of creating a high-resolution (HR) image, given one or multiple low-resolution (LR) images as input. The SR process aims at adding to the LR input(s) new plausible high frequency details, to a greater extent than traditional interpolation methods. We mostly focused on the single-image problem, where only a single LR image is available. We have adopted the example-based framework on one hand and the sparse approximation framework on the other hand.

In the example-based framework, the relation between the LR and HR image spaces is modeled with the help of pairs of small "examples", i.e. texture patches. Each example pair consists of a LR patch and its HR version that also includes high-frequency details; the pairs of patches form a dictionary of patches. For each patch of the LR input image, one or several similar patches are found in the dictionary, by performing a nearest neighbor search. The corresponding HR patches in the dictionary are then combined to form a HR output patch; and finally all the reconstructed HR patches are re-assembled to build the superresolved image. In this procedure, one important aspect is how the dictionary of patches is built. At this regard, two choices are possible: an external dictionary, formed by sampling HR and LR patches from external training images; and an internal dictionary, where the LR/HR patch correspondences are learned by putting in relation directly the input image and scaled versions of it. The advantage of having an external dictionary is that it is built in advance: this leads to a reduction of the computational time, whereas in the internal case the dictionary is generated online at each run of the algorithm. However, external dictionaries have a considerable drawback: they are fixed and so non-adapted to the input image. To be able to satisfactorily process any input image, we need then to include in the dictionary a large variety of patch correspondences, leading to a high computational time. In 2013, external dictionaries have been designed to bridge the gap between external and internal dictionary based methods.

In 2014 instead, we proposed a novel SR method for internal dictionaries [16]. The internal dictionary contains pair of LR/HR patches taken from the image to be processed and is by construction well adapted to the data. However, its size is limited since it results from the sampling of a single image. This leads to an undersampling of the LR space and even more of the HR space. To overcome this problem, state of the art methods select, for each input LR patch, a local neighborhood, learn the local geometry of this neighborhood, and apply it in the HR domain. Therefore, an underlying hypothesis is that the local neighborhoods in the LR and HR domain are similar. To avoid this hypothesis, we employ a regression-based method to directly map LR input patches into their related HR output patches. To make this regression more robust, first the LR patches have been first oversampled (by a bicubic interpolation) such that LR and HR spaces have the same dimension, and second a Tikhonov regularization has been added. When compared to other state-of-the-art algorithms, our proposed algorithm shows the best performance, both in terms of objective metrics and subjective visual results. As for the former, it presents considerable gains in PSNR and SSIM values. When observing the super-resolved images, also, it turns out to be the most capable in producing fine artifact-free HR details.

### 6.2.3. Image super-resolution in a sparse and manifold learning framework

**Participants:** Julio Cesar Ferreira, Christine Guillemot, Olivier Le Meur, Elif Vural.

The problem of image super-resolution has also been addressed in a sparse approximation framework. This led to a novel algorithm based on sparse representations in which a structure tensor-based regularization has been introduced [29]. The relative discrepancy between the two eigenvalues of the structure tensor is an indicator of the degree of anisotropy of the gradient in a region of the image. The eigenvalues and eigenvectors of the structure tensor are used to compute, for each pixel belonging to a salient edge, a stream line in the direction perpendicular to the edge (given by the eigenvector corresponding to the highest eigenvalue of the structure tensor). The saliency of an edge is given by the S-norm of the highest eigenvalue. An energy term dealing with the sharpness of edges is then computed and used as a regularization constraint to modify the current estimated high resolution image inside the Iterative Shrinkage Thresholding algorithm. This extra constraint forces the value of the current pixel along the stream line to be as close as possible to pixel values having lowest saliency. The resulting single-image algorithm, called Sharper Edges based Adaptive Sparse Domain Selection (SE-ASDS) allows sharpening edges and reducing the ringing artefacts compared to existing methods. This is illustrated in Fig.3

*Figure 3. Comparison of SR results ($\times 3$). (a) LR image;(b) Nearest-neighbor; (c) Sparse method without structure-based regularization; (d) SE-ASDS results. (e) Comparison between (c) and (d) on patches: edges of (d) are more contrasted than (c).*

In the previous method, the dictionaries used for the sparse approximation method are defined as a union of PCA basis learned on clusters of patches of the input image. The clusters are constructed using the classical k-means algorithm with patch distances computed with the Euclidean distance. This study is being pursued by assuming manifold models for the patches of the input images. A method using graph-based clustering has then been used for clustering patches on the manifold, and this method has been extended to cope with the out-of-sample problem. Dedicated dictionary learning methods are currently under development to have dictionaries best adapted to the manifold structure.

## 6.3. Representation and compression of large volumes of visual data

Sparse representations, data dimensionality reduction, compression, scalability, perceptual coding, rate-distortion theory

### 6.3.1. Manifold learning and low dimensional embedding for classification

**Participants:** Christine Guillemot, Elif Vural.

Typical supervised classifiers such as SVM are designed for generic data types and do not make any particular assumption about the geometric structure of data, while data samples have an intrinsically low-dimensional structure in many data analysis applications. Recently, many supervised manifold learning methods have been proposed in order to take the low-dimensional structure of data into account when learning a classifier. Unlike unsupervised manifold learning methods which only take the geometric structure of data samples into account when learning a low-dimensional representation, supervised manifold learning methods learn an embedding that not only preserves the manifold structure in each class, but also enhances the separation between different classes.

An important factor that influences the performance of classification is the separability of different classes in the computed embedding. We thus do a theoretical analysis of separability of data representations given by supervised manifold learning. In particular, we focus on the nonlinear supervised extensions of the Laplacian eigenmaps algorithm and examine the linear separation between different classes in the learned embedding. We first consider a setting with two classes and show that the two classes become linearly separable even with a one-dimensional embedding. We characterize the linear separation in terms of the data graph properties such as edge weights, diameter, and volume and some algorithm parameters. We then extend these results to a setting with multiple classes, where the classes are assumed to be categorizable into a few groups with high intra-group affinities. We show that, if the graph is such that the inter-group graph weights are sufficiently small, the learned embedding becomes linearly separable at a dimension that is proportional to the number of groups. These theoretical findings are also confirmed by experimentation on synthetic data sets and image data.

Next, we consider the problem of out-of-sample generalizations for manifold learning. Most manifold learning methods compute an embedding in a pointwise manner, i.e., data coordinates in the learned domain are computed only for the initially available training data. The generalization of the embedding to novel data samples is an important problem, especially in classification problems. Previous works for out-of-sample generalizations are designed for unsupervised methods. We study the problem for the particular application of data classification and propose an algorithm to compute a continuous function from the original data space to the low-dimensional space of embedding. In particular, we construct an interpolation function in the form of a radial basis function that maps input points as close as possible to their projections onto the manifolds of their own class. Experimental results show that the proposed method gives promising results in the classification of low-dimensional image data such as face images.

### 6.3.2. Dictionary learning for sparse coding and classification of satellite images

**Participants:** Jeremy Aghaei Mazaheri, Christine Guillemot, Claude Labit.

In the context of the national partnership Inria-Astrium, we explore novel methods to encode images captured by a geostationary satellite. These pictures have to be compressed on-board before being sent to earth. Each picture has a high resolution and so the rate without compression is very high (about 70 Gbits/sec) and the goal is to achieve a rate after compression of 600 Mbits/sec, that is a compression ratio higher than 100. On earth, the pictures are decompressed with a high reconstruction quality and visualized by photo-interpreters. The goal of the study is to design novel transforms based on sparse representations and learned dictionnaries for satellite images.

We have developed methods for learning adaptive tree-structured dictionaries. Each dictionary in the structure is learned on a subset of residuals from the previous level, with the K-SVD algorithm. The tree structure offers better rate-distortion performance than a "flat" dictionary learned with K-SVD, especially when only a few atoms are selected among the first levels of the tree. The tree-structured dictionary allows efficient coding of the indices of the selected atoms. Besides coding, these structured dictionaries turn out to be useful tools for

MTF (Modulation Transfer Function) estimation and supervised classification. The MTF estimation consists in estimating the MTF of the instrument used to take this picture. The learned structured dictionaries are currently studied to perform supervised classification in a context of scene recognition in satellite images. In that case, dictionaries should be learned for specific scenes. Then, patches (around each pixel) of a test picture to classify are decomposed over the different dictionaries to determine for each pixel the dictionary giving the best approximation and thus the corresponding class. A graph-cut algorithm can be applied to smooth the classification results. We are currently trying to learn more discriminant dictionaries for this specific application. For that purpose, the objective function to minimize to learn the dictionaries should not only be reconstructive, but also discriminative.

### *6.3.3. Adaptive clustering with Kohonen self-organizing maps for second-order prediction*

**Participants:** Christine Guillemot, Bihong Huang.

The High Efficiency Video Coding standard (HEVC) supports a total of 35 intra prediction modes which aim at reducing spatial redundancy by exploiting pixel correlation within a local neighborhood. However the correlation remains in the residual signals of intra prediction, leading to some high energy prediction residuals. In 2014, we have studied several methods to exploit remaining correlation in residual domain after intra prediction. The method uses vector quantization with codebooks learned and dedicated to the different prediction modes in order to model the directional characteristics of the residual signals. The best matching code vector is found in a rate-distortion optimization sense. Finally, the index of the best matching code vector is sent to the decoder and the vector quantization error, the difference between the intra residual vector and the best matching code vector, is processed by the conventional operations of transform, scalar quantization and entropy coding. In a first approach, the codebooks are learned using the k-means algorithm. The learning algorithm proceeds in two passes so that the training set of residual vectors corresponds to the case where the vector quantization is the best mode in rate-distortion sense for the second-order prediction. It has been observed that the codebooks learned for different Quantization Parameters (QP) are very similar, leading eventually to QP-independent codebooks. A second method is being developed using clustering with Kohonen self-organizing maps in the codebook learning stage.

### *6.3.4. HDR video compression*

**Participants:** Christine Guillemot, Mikael Le Pendu.

High Dynamic Range (HDR) images contain more intensity levels than traditional image formats. Instead of 8 or 10 bit integers, floating point values requiring much higher precision are used to represent the pixel data, leading to new compression challenges. In collaboration with Technicolor, we have developed a method for converting the floating point RGB values to high bit depth integers with an approximate logarithmic encoding that is reversible without loss. This bit depth reduction is performed adaptively depending on the minimum and maximum values which characterize the dynamic of the data. A 50% rate saving has been obtained at high bitrates compared to the well-known adaptive LogLuv transform [33]. A reversible tone mapping-operator (TMO) has also been designed for efficient compression of High Dynamic Range (HDR) images using a Low Dynamic Range (LDR) encoder. Based on a statistical model of the HDR compression scheme and assumptions on the rate of the encoded LDR image, a closed form solution has been derived for the optimal tone curve in a rate-distortion sense [34].

### *6.3.5. HEVC-based UHD video coding optimization*

**Participants:** Nicolas Dhollande, Christine Guillemot, Olivier Le Meur.

The HEVC (High Efficiency Video Coding) standard brings the necessary quality versus rate performance for efficient transmission of Ultra High Definition formats (UHD). However, one of the remaining barriers to its adoption for UHD content is the high encoding complexity. We address the problem of HEVC encoding complexity reduction by proposing a strategy to infer UHD coding modes and quadtree from those optimized on the lower (HD) resolution version of the input video. A speed-up by a factor of 3 is achieved compared to directly encoding the UHD format at the expense of a limited PSNR-rate loss [28]. Another method which is still under investigation is to extract from the input video sequence a number of low-level features for

adapting the coding decision such as the decomposition of the quadtree. The low-level features are related to gradient-based statistics, structure tensors statistics or entropy etc.

## 6.4. Distributed processing and robust communication

Information theory, stochastic modelling, robust detection, maximum likelihood estimation, generalized likelihood ratio test, error and erasure resilient coding and decoding, multiple description coding, Slepian-Wolf coding, Wyner-Ziv coding, information theory, MAC channels

### *6.4.1. Universal distributed source coding*

**Participant:** Aline Roumy.

In 2012, we started a new collaboration with Michel Kieffer and Elsa Dupraz (Supelec, L2S) on universal distributed source coding. Distributed source coding (DSC) refers to the problem where several correlated sources need to be compressed without any cooperation at the encoders. Decoding is however performed jointly. This problem arises in sensor networks but also in video compression techniques, where the successive frames are seen as distributed such that the correlation between the frames is not directly used at the encoder. Traditional approaches for DSC (from an information theoretical but also practical point of view) assume that the joint distribution of the sources is perfectly known. Since this assumption is not satisfied in practice, a way to get around this is to use a feedback channel (from the decoder to the encoder), that can trigger the encoder. Instead, we consider universal distributed source coding, where the joint source distribution is unknown.

More precisely, we considered the problem of compressing one source, while a second source, called side information, is available at the decoder. Further, we assumed that the conditional distribution of the side information given the source is unknown at both encoder and decoder. First, we proposed in [18] four uncertainty models for this conditional distribution, and derived the information theoretical bounds. These models differ through the (partial) knowledge on the distribution the user has. This partial knowledge includes the variation speed (slow/fast), the set of possible distributions, and eventually, some a priori distribution on the class of distributions. A complete coding scheme has also been proposed that works well for any distribution in the class. At the encoder, the proposed scheme encompasses the determination of the coding rate and the design of the encoding process. These determinations directly result from the information-theoretical compression bounds. Then a novel decoder is proposed that jointly estimate the source symbols and the conditional distribution. As the proposed decoder is based on the Expectation-Maximization algorithm, which is very sensitive to initialization, we also propose a method to produce first a coarse estimate of the distribution. The proposed scheme avoids the use of a feedback channel or the transmission of a learning sequence, which both result in a rate increase at finite length. Moreover, the proposed algorithm use non-binary LDPC codes, such that the usual binarization of the source, which induce compression inefficiency, can be avoided.

### *6.4.2. Rate Distortion analysis of Compressed sensing and distributed Compressed sensing*

**Participant:** Aline Roumy.

In collaboration with Enrico Magli and Giulio Coluccia (Polito, Torino, Italy), we studied Compressed sensing as a communication tool. Compressed sensing (CS) is an efficient acquisition scheme, where the data are projected onto a randomly chosen subspace to achieve data dimensionality reduction. The projected data are called measurements. The reconstruction is performed from these measurements, by solving underdetermined linear systems under a sparsity a priori constraint. However, the obtained measurements are reals, and therefore require an infinite precision representation. Therefore, using CS as a compression tool (in the information theoretical sense), requires to determine the trade-off between the rate necessary to encode the measurements and the distortion obtained on the data. In [17], we derive the rate-distortion (RD) function of CS and distributed CS, under the assumption that the sparsity support is perfectly known at the decoder. This provides a lower bound for any practical reconstruction algorithm.

The proof technique developed in [17] has application beyond information theory. It also provides novel analyses of CS reconstruction algorithms [27]. Classical performance analysis of reconstruction algorithms, rely on parameters that are difficult to compute (RIP, coherence of the measurement matrix), for which bounds are used. Instead, we derive exact characterization, by performing either averaged (over the measurement matrix) or asymptotic (in the size of the data) analysis.

**SUMO Project-Team**

# 6. New Results

## 6.1. Highlights of the Year

We started our first industrial collaboration "Project P22" with Alstom Transport, in the context of a common laboratory between Inria and Alstom. The project started in March 2014 and tackles robustness issues and regulation in urban train systems. The second phase of the project will start in march 2015, for a duration of three years. Most of the researchers of Sumo are involved in this project.

## 6.2. Control and enforcement

### *6.2.1. Runtime enforcement of timed properties*

**Participants:** Thierry Jéron, Hervé Marchand, Srinivas Pinisetty.

Runtime enforcement is a powerful technique to ensure that a running system satisfies some desired properties. Using an enforcement monitor, an (untrustworthy) input execution (in the form of a sequence of events) is modified into an output sequence that complies with a property. Over the last decade, runtime enforcement has been mainly studied in the context of untimed properties. The contributions [26] and [34] deal with runtime enforcement of timed properties by revisiting the founda-tions of runtime enforcement when time between events matters. We propose a new enforcement paradigm where enforcement mechanisms are time retardants: to produce a correct output sequence, additional delays are introduced between the events of the input sequence. We consider runtime enforcement of any regular timed property defined by a timed automaton. We prove the correctness of enforcement mechanisms and prove that they enjoy two usually expected features, revisited here in the context of timed properties. The first one is soundness meaning that the output sequences (eventually) satisfy the required property. The second one is transparency, meaning that input sequences are modified in a minimal way. We also introduce two new features, i) physical constraints that describe how a time retardant is physically constrained when delaying a sequence of timed events, and ii) optimality, meaning that output sequences are produced as soon as possible. To facilitate the adoption and implementation of enforcement mechanisms, we describe them at several complementary abstraction levels. Our enforcement mechanisms have been implemented and our experimental results demonstrate the feasibility of runtime enforcement in a timed context and the effectiveness of the mechanisms. Finally, in [33], we considered more practical applications. Indeed, in network security, RE monitors can detect and prevent Denial-of-Service attacks. In resource allocation, RE monitors can ensure fairness. Specifications in these domains express data-constraints over the received events where the timing between events matters. To formalize these requirements, we introduce Parameterized Timed Automata with Variables (PTAVs), an extension of Timed Automata (TAs) with internal and external variables. We then extend enforcement for TAs to enforcement for PTAVs for safety properties. We model requirements from the considered application domains and show how enforcement monitors can ensure system correctness w.r.t. these requirements.

### *6.2.2. Enforcing opacity*

**Participant:** Hervé Marchand.

In [22], we have been interested in enforcing opacity of regular predicates on modal transition systems. Intuitively, a labelled transition system $\mathcal{T}$ partially observed by an attacker, and a regular predicate $S$ over the runs of $\mathcal{T}$, enforcing opacity of the secret $S$ in $\mathcal{T}$ means computing a supervisory controller $K$ such that an attacker who observes a run of the controlled system $K \setminus \mathcal{T}$ cannot ascertain that the trace of this run belongs to $S$ based on the knowledge of $\mathcal{T}$ and $K$. We lift the problem from a single labelled transition system $\mathcal{T}$ to the class of all labelled transition systems specified by a *Modal Transition System*$\mathcal{M}$. The lifted problem is to compute the maximally permissive controller $K$ such that $S$ is opaque in $K/\mathcal{T}$ for every labelled transition system $\mathcal{T}$ which is a model of $\mathcal{M}$. The situations of the attacker and of the controller are asymmetric: at run time, the attacker may fully know $\mathcal{T}$ and $K$ whereas the controller knows only $\mathcal{M}$ and the sequence of actions executed so far by the unknown $\mathcal{T}$.

In [23], we provided a different solution by enforcing and validate ay runtime various notion of opacity. More specically, we studied how we can model-check, verify and enforce at system runtime, several levels of opacity. Besides existing notions of opacity, we also introduce K-step strong opacity, a more practical notion of opacity that provides a stronger level of confidentiality.

### *6.2.3. Discrete Controller Synthesis for Infinite State Systems with ReaX*

**Participants:** Nicolas Berthier, Hervé Marchand.

This year, we investigated the control of infinite reactive synchronous systems modeled by arithmetic symbolic transition systems for safety properties handling numerical variable. We provide effective algorithms allowing to solve the safety control problem, and report on experiments based on ReaX, our tool implementing these algorithms [28].

## 6.3. Model expressivity and quantitative verification

### *6.3.1. Diagnosis*

**Participants:** Nathalie Bertrand, Sébastien Chédor, Éric Fabre, Loïc Hélouët, Blaise Genest, Hervé Marchand, Christophe Morvan.

Diagnosis of a system consists in providing explanations to a supervisor from a partial observation of the system and a model of possible executions. This year, we have extended results on diagnosis algorithm from scenarios. Systems are modeled using High-level Message Sequence Charts (HMSCs), and the diagnosis is given as a new HMSC, which behaviors are all explanations of the partial observation. The results published this year are first an offline centralized diagnosis algorithm (a single process in a network collects an observation, and emits a diagnosis) that has then been extended to a decentralized version of this algorithm. This allows us to give a complete diagnosis framework for infinite state systems, with a strong emphasis on concurrency and causal ordering in behaviors. HMSC-based diagnosis showed nice properties w.r.t. compositionality. We have also considered solutions for online diagnosis from scenarios, but came to the conclusion that online solutions are memory consuming, and need too many restrictions to run with finite memory. The last contribution of this work is an application of diagnosis techniques to anomaly detection, that is a comparison of observation of the system with a model of usual behaviors to detect security attacks. This work has been published this year [24].

In [21] we have been interested in the analysis of discrete event systems under partial observation which is an important topic, with major applications such as the detection of information flow and the diagnosis of faulty behaviors. These questions have, mostly, not been addressed for classical models of recursive systems, such as pushdown systems and recursive state machines. In this paper, we consider recursive tile systems, which are recursive infinite systems generated by a finite collection of finite tiles, a simplified variant of deterministic graph grammars (slightly more general than pushdown systems). Since these systems are infinite-state in general powerset constructions for monitoring do not always apply. We exhibit computable conditions on recursive tile systems and present non-trivial constructions that yield effective computation of the monitors.We apply these results to the classic problems of state-based opacity and diagnosability (off-line verification of opacity and diagnosability, and also run-time monitoring of these properties). For a decidable subclass of recursive tile systems, we also establish the decidability of the problems of state-based opacity and diagnosability.

In discrete event systems prone to unobservable faults, a diagnoser must eventually detect fault occurrences. The diagnosability problem consists in deciding whether such a diagnoser exists. We laid the foundations of diagnosis and predicatability for probabilistic systems represented by partially observed Markov chains (denoted pLTS) [32]. In particular, we studied different specifications of diagnosability and establish their relations both in finite and infinite pLTS. Then we analyzed the complexity of the diagnosability problem for finite pLTS: we showed that the polynomial time procedure proposed earlier is erroneous and that in fact for all considered specifications, the problem is PSPACE-complete. We also established tight bounds for the size of diagnosers. Afterwards we considered the dual notion of predictability which consists in predicting that in a safe run, fault will eventually occur. Predictability is easier than diagnosability: it is NLOGSPACE-complete. Yet the predictor synthesis is as hard as the diagnoser synthesis.

When a system is not diagnosable, the active diagnosis problem consists in controlling the system in order to ensure its diagnosability. In the same probabilistic setting, the active diagnosis problem consists in deciding whether there exists some observation-based strategy that makes the system diagnosable with probability one. We proved that this problem is EXPTIME-complete, and that the active diagnosis strategies are belief-based. The *safe* active diagnosis problem is similar, but aims at enforcing diagnosability while preserving a positive probability to non faulty runs, i.e. without enforcing the occurrence of a fault. We prove that this problem requires non belief-based strategies, and that it is undecidable. However, it belongs to NEXPTIME when restricted to belief-based strategies. Our work also refines the decidability/undecidability frontier for verification problems on partially observed Markov decision processes [30].

### 6.3.2. Probabilistic model checking

**Participants:** Nathalie Bertrand, Blaise Genest, Paulin Fournier.

In [16], we considered the verification of Markov chains against properties talking about distributions of probabilities. Even though a Markov chain is a very simple formalism, by discretizing in a finite number of classes the space of distributions through some symbolics, we proved that the language of trajectories of distribution (one for each initial distribution) is not regular in general, even with 3 states. We then proposed a parametrized algorithm which approximate what happens to infinity, such that each symbolic block in the approximate language is at most $\epsilon$ away from the concrete distribution.

Parameterized verification aims at validating a model of a system irrespective of the value of a parameter. This year, we studied verification problems for a model of network with the following characteristics: the number of entities is parametric, communication is performed through broadcast with adjacent neighbors, entities can change their internal state probabilistically and reconfiguration of the communication topology can happen at any time. The semantics of such a model is given in term of an infinite state system with both non deterministic and probabilistic choices. We are interested in qualitative problems like whether there exists an initial topology and a resolution of the non determinism such that a configuration exhibiting an error state is almost surely reached. We showed in [44] that all the qualitative reachability problems are decidable and some proofs are based on solving a 2 player game played on the graphs of a reconfigurable network with broadcast with parity and safety objectives.

On a different topic, we considered a control problem for stochastic sytems specified by timed automata with distributions over delays. In [29] we considered reachability objectives on such decision stochastic timed automata (DSTA). Given a reachability objective, the value 1 problem asks whether a target can be reached with probability arbitrarily close to 1. Simple examples show that the value can be 1 and yet no strategy ensures reaching the target with probability 1. In this paper, we prove that, the value 1 problem is decidable for single clock DSTA by non-trivial reduction to a simple almost-sure reachability problem on a finite Markov decision process. The $\varepsilon$-optimal strategies are involved: the precise probability distributions, even if they do not change the winning nature of a state, impact the timings at which $\varepsilon$-optimal strategies must change their decisions, and more surprisingly these timings cannot be chosen uniformly over the set of regions.

### 6.3.3. Distributed timed systems

**Participants:** Blaise Genest, Loïc Hélouët.

We have proposed and considered properties of a new timed variant of Petri nets [42], namely Timed Petri Nets with Urgency, that extend Timed Petri Nets with the main features of TPNs. Time Petri Nets (TPN) [52] and Timed Petri Nets [45] are two incomparable classes of concurrent models with timing constraints: urgency cannot be expressed using Timed Petri Nets, while TPNs can only keep track of a bounded number of continuous values (clocks). The work performed this year provides up to-our-knowledge the first decidability results for Petri Net variants combining time, urgency and unbounded places. We have obtained decidability of control-state reachability for the subclass of Timed Petri Nets with Urgency where urgency constraints can only be used on bounded places. By restricting this class to use a finite number of clocks, we have shows decidability of (marking) reachability. Formally, this class corresponds to TPNs under a new, yet natural, timed semantics where urgency constraints are restricted to bounded places. Further, under their original semantics, reachability for a more restricted class of TPNs is decidable.

### 6.3.4. Test Generation from Recursive Tile Systems

**Participants:** Sébastien Chédor, Christophe Morvan, Thierry Jéron.

In [20] we explore the generation of conformance test cases for *Recursive Tile Systems* in the framework of the classical **ioco** testing theory. The RTS model allows the description of reactive systems with recursion, and is very similar to other models like Pushdown Automata, Hyperedge Replacement Grammars or Recursive State Machines. Test generation for this kind of infinite state labelled transition systems is seldom explored in the literature. The first part presents an off-line test generation algorithm for *Weighted* RTSs, a determinizable subclass of RTSs, and the second one, an on-line test generation algorithm for the full RTS model. Both algorithms use test purposes to guide test selection through targeted behaviours. Additionally, essential properties relating verdicts produced by generated test cases with both the soundness with respect to the specification, and the precision with respect to a test purpose, are proved.

## 6.4. Management of large distributed systems

### 6.4.1. Distributed optimal planning

**Participant:** Éric Fabre.

Planning problems consist in organizing actions in a system in order to reach one of some target states. The actions consume and produce resources, can of course take place concurrently, and may have costs. We have a collection of results addressing this problem in the setting of distributed systems. This takes the shape of a network of components, each one holding private actions operating over its own resources, and shared/synchronized actions that can only occur in agreement with its neighbors. The goal is to design in a distributed manner a tuple of local plans, one per component, such that their combination forms a consistent global plan of minimal cost.

Our previous solutions to this problem modeled components as weighted automata. In collaboration with Loig Jezequel (TU Munich) and Victor Khomenko (Univ. of Newcastle), we have extended this approach to the case of components modeled as safe Petri nets [50]. This allows one to benefit from the internal concurrency of actions within a component. Benchmarks have shown that this method can lead to significant time reductions to find feasible plans, in good cases. In the least favorable cases, performances are comparable to those obtained with components modeled as automata. The method does not apply to all situations however, as computations require to perform $\epsilon$-reductions on Petri-nets (our work also contains a contribution to this difficult question). This work has been accepted by the ACM Transactions in Embedded Computing Systems, to appear in 2015.

## 6.5. Data driven systems

### 6.5.1. Web services

**Participants:** Blaise Genest, Loïc Hélouët.

This year, we considered transactional properties (ACID) for web services. In particular, we focused on the atomicity (A of ACID) property, obtained in case of a failure inside an atomic block through compensation of the executed actions of the block. To do so, logs need to be kept. We were interested in maintaining the maximal amount of privacy. We proposed modular algorithms [19] which maintain privacy between modules, with minimal information shared among modules, both in the logging and the compensation phases. Furthermore, each module logs a small number of information, such that the sum of all actions logged is guaranteed minimal. Last, modularity allows fast algorithms, as they need to consider only what happens in the module itself, and not the exact structure of its parent module nor of its sub-modules.

We also published results on our model of sessions systems [27]. This models allows for the modeling of distributed web-based systems that are running an arbitrary number of transactions among arbitrarily many participants. We have shown how simple restrictions can guarantee decidability of simple coverability properties, and then be used to detect violation of buisness rules such as conflict of interest, or a more complex property called the chinese wall.

We are currently considering new models that manage at the same time explicit workflows and structured data. This model can be seen as a combination of AXML [46] and Petri nets.

### *6.5.2. An Artifact-centric Process Model*

**Participants:** Éric Badouel, Loïc Hélouët, Christophe Morvan.

In [37] we present a purely declarative approach to artifact-centric case management systems, and a decentralization scheme for this model. Each case is presented as a tree-like structure; nodes bear information that combines data and computations. Each node belongs to a given stakeholder, and semantic rules govern the evolution of the tree structure, as well as how data values derive from information stemming from the context of the node. Stakeholders communicate through asynchronous message passing without shared memory, enabling convenient distribution.

**TACOMA Team**

# 6. New Results

## 6.1. Self-describing objects and tangible data structures

**Participants:** Nebil Ben Mabrouk, Paul Couderc [contact], Arnab Sinha.

A development in the line of the coupled objects principles are self-describing objects. While previous works enabled integrity checking over a set of physical objects, these mechanisms were limited in two aspects: expressiveness and autonomy. More precisely, coupled objects support the detection of special conditions (such as a missing element), but not the characterization of these conditions (such as describing the problem, identifying the missing element). Moreover, this compromises the autonomous feature of coupled objects, which would depend on external systems for analyzing these special conditions. Self-describing objects are an attempt to overcome these limitations, and to broaden the application perspectives of autonomous RFID systems.

The principle is to implement distributed data structure over a set of RFID tags, enabling a complex object (made of various parts) or a set of objects belonging to a given logical group to "ÃÂÃÂself-describe" itself and the relation between the various physical elements. Some applications examples includes waste management, assembling and repair assistance, prevention of hazards in situations where various products / materials are combined etc. The key property of self-describing objects is, like for coupled objects, that the vital data are self-hosted by the physical element themselves (typically in RFID chips), not an external infrastructure like most RFID systems. This property provides the same advantages as in coupled objects, namely high scalability, easy deployment (no interoperability dependence/interference), and limited risk for privacy.

However, given the extreme storage limitation of RFID chips, designing such systems is difficult:

- Data structures must be very frugal in terms of space requirements, both for the structure and for the coding.
- Data structures must be robust and able to survive missing or corrupted elements if we want to ensure the self-describing property for a damaged or incorrect object.

In the context of RFID system, the resiliency property of such data structures enables new information architecture and autonomous (offline) operation, which is very important for some RFID applications. On this topic, a generic graph structure applicable to RFID systems for supporting self-describing objects is proposed in Arnab Sinha's thesis document [1], and was published in [4].

## 6.2. Pervasive support for RFIDs

**Participants:** Nebil Ben Mabrouk, Paul Couderc [contact].

In situations where we have to read large collection of objects of various types, the performance is difficult to predict but may still be adequate for a given application. For example, some application can tolerate missing some tags, provided that miss read probability could be characterized. In some cases, read reliability could be improved using mechanical approaches, such as introducing movements in objects or antenna to introduce radio diversity during read. Finally, distributed data structure can be used over a set of tags to be used to mitigate the impact of mis-read (by using data redundancy) and to help the reading protocol by integrating hints about the tag set collection being read.

Our objective here is to study extensively by experimentation the behaviour of existing RFID solutions in the context of uncontrolled environment (meaning, random placement of tags on objects mixing various materials) in order to characterize their real-world performance regarding the parameters of such as tags numbers, density, frequencies, reader antenna design, dynamicity of objects (movements), etc. From these experimentations, we would like to identify the conditions that are favorable to acceptable performance, and the way where there are hopes of improvement with specific design for these difficult environments. These results should also allow improving the performance: high level integrity checks can guide low level operations by determining whether inventories are complete or not. This cross layer strategy can enable faster are more efficient inventory protocols.

An important milestone was completed in 2014, with the implementation of an experiment test bed in order to support the experiment campaign. This task involved a significant development and engineering effort. This testbed is currently deployed at the IETR (http://www.ietr.fr) building, and features a multi-axis mobile RFID antenna system driven by a software platform.

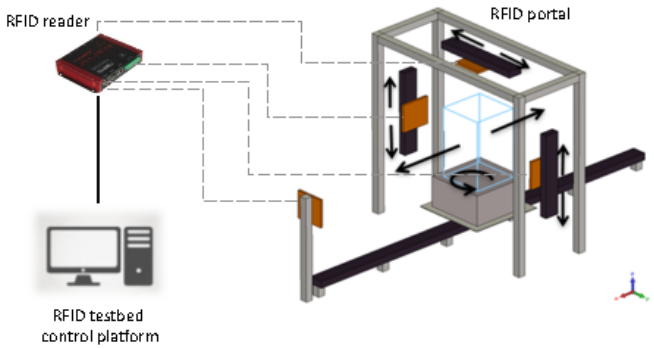


*Figure 1. RFID testbed*

This system allows both interactive testing as well as long running experiments of RFID reading protocols. The software platform was designed to allow 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). Given this flexibility, this platform should be able to reproduce most of the situations found in real applications. In particular, it can be used to design custom reading set up optimized for various RFID portal applications [3].

## 6.3. Context-aware dynamic Smart Home Platform

**Participants:** Andrey Boytsov, Aurélien Richez, Yoann Maurel, Frédéric Weis [contact].

Tacoma group is focussed on the conception and implementation of innovative services for the Smart Home. The range of considered services is broad : from "optimizing the energy consumption" to "helping users to find their way in a building". To provide such services, automation based on pre-set scenarios is ineffective: human behavior is hardly predictable and application should be able to adapt their behavior at runtime depending on the context. We focused on recognizing user's activities to adapt applications behaviours.

Building efficient and accurate context awareness was and is still a great challenge but we proved, through the use of dedicated algorithms and a layered architecture that it is achievable when the targeted Home is known - due to the specific and non automated calibration process we used. Among all the available theories, we decided to use the Belief Function Theory (BFT) [8] [9] as it allows to express uncertainty

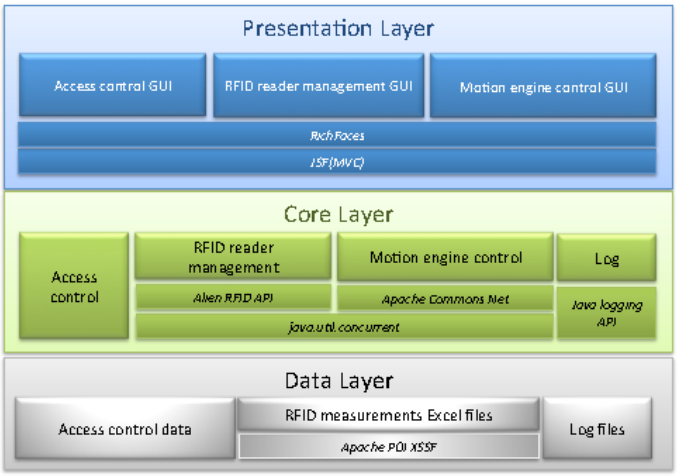


*Figure 2. Software architecture of the RFID testbed*

and imprecision. Although these results are very promising, great challenges still lied in (i) the support of the dynamic reconfiguration to face evolving hardware or software conditions and (ii) the deployment and the configuration of the layered architecture and sensors to allow the use of our approach in unknown environments.

One of our goals is to build a pervasive platform with constrained performance and cost [7]. The cost is particularly critical for sensors and actuators: we choose to limit our scope to inexpensive and non-invasive sensors *i.e.* no video camera. This past months, Tacoma has been working on the conception and implementation of a Smart Home Platform based on earlier prototypes inherited from ACES team. The prototypes were implemented as an hard-to-maintain monolithic code. The code also suffered from a lot of redundancy. More importantly the platform hardly supported dynamism and provided no support for reconfiguration and adaptation at runtime. With this in mind, during the re-writing of the platform the emphasis has been placed on the following aspects:

- supporting the dynamic discovery of heterogeneous sensors;
- enabling the dynamic deployment of applications at runtime ;
- enabling context-awareness by providing contextual information to these applications;
- enhancing the separation of concerns and code-reuse.

Our goal is to design and build a platform that is:

- **evolutive**: the Home environment is ever-changing and thus it is important to allow users to add new sensors or new services dynamically at runtime. It is also mandatory to recalibrate the sensors to face the change in the Home. This is mainly why we based our platform on OSGi;
- **maintainable and administrable**: we raised the maintainability by using a modular approach using C-modules or iPOJO components; the platform is itself modular to achieve a good separation of concerns (e.g., communication, module loading, discovery). We also built in-production monitoring interfaces that provides information on the belief functions that are used, the fusion process and the sensors values;
- **easy to configure**: alleviating the complexity of the platform configuration and maintenance is a prerequisite for the adoption of Smart-Home environments by consumers. Currently the BFT theories requires a huge calibration process. We focussed our efforts on the semi-automated building of mass functions, required by the theory, that have to be provided by each sensor.

### 6.3.1. Towards dynamism using OSGI

The development of our initial platform in C proved to be costly and hard to maintain. The dynamism is hard to achieved with a low-level language and requires an heavy development process. This led the team to investigate the use of OSGi as a based for our execution platform. OSGi is the specification of an execution framework developed on top of Java. It relies on the Java's dynamic features (dynamic and on demand class loading through class loaders) to provide a coarse-grained level of modularity. This choice was reinforced by our collaboration with the Adele team (LIG Laboratory in Grenoble). This team is using OSGi as a core for building Smart Home applications. Using OSGi would ease collaboration and code sharing.

One main concerns regarding the use of Java was the limited performances of the targeted hardware (raspberry pi). The Belief Function Theory (BFT) requires heavy computations and the embedded CPU could have been the bottleneck. Moreover, the JVM supported by the raspberry pi is limited compared to standard JVM. As a preliminary study, we choose to implement the core of the BFT library in Java and to compare the performances with the C implementation. Unexpectedly the Java implementation performed better than the C implementation in most of the case. This can be explained by three factors. First, the BFT theory is tedious to implement in low-level language. The C-implementation could probably be optimized but this will lower the readability of the source code and impact the maintainability. Second and conversely, using Java raised the code readability and allowed us to performed some optimization. Third, the JIT (Just In Time) compiler provided by the VM have been improved these past years and the optimization performed by the VM are sufficient to bring on par performances with the C implementation. As, the performances of the C platform were largely sufficient, this preliminary phase validated our decision to switch to OSGi.

### 6.3.2. Automated configuration of sensors

A previous defended in the group in december 2013 has shown promising results applying the BFT theory to the Smart Home Domain. It is currently possible to collect sensor values and extract belief functions from them. The platform can then extract a context from the belief functions and offer services to the user depending on what is happening. For instance, the user may be notified of an open window when he leaves the house.

The transition between a raw sensor value and a belief function is made through the use of a belief model which maps a sensor value to a belief function. The belief model is provided to the platform by us and a component is in charge of transforming a sensor value in a belief function. The fine tuning of a model can be a tedious task. It must be done by a specialist who understands the belief function theory and knows the behavior of the sensors. The model is often built iteratively by experimenting. This may take several hours or days.

Ideally, the calibration of the model should be as automatic as possible (few interaction with the user during calibration). The person setting up the sensors should not have to understand the belief function theory. The group is currently studying the possible use of clustering and classifications algorithm in order to ease the calibration of sensors. Yoann Maurel and Frédéric Weis supervised a project with a group of ENS student on this subject. The goal is to generate our belief model from a training set of sensor data. We mainly focus on two algorithms: k-nearest neighbors (KNN) and overlapping k-mean (OKM). A first experimentation with KNN and motion sensors showed that this algorithm is promising. We used a training data set to compute the presence belief model. We acquired a first set of data with someone present in the experimentation room and a second data set with nobody in the room, which gives us a labelled data set.

## 6.4. Towards Metamorphic Housing: the on-demand room

**Participant:** Michele Dominici [contact].

This research activity is supported by Fondation Rennes 1 through the chair "Smart Home and Innovation", since January 2014. During the first year, we focused on identifying the needs of the industrial partners and public authorities that fund the chair.

This activity is centered on the concept of metamorphic housing (see section 4.2 ). During this year, we introduced a solution of metamorphic housing addressing the goals of saving space and energy in an apartment building, while preserving residents' comfort: the on-demand room. It consists in a space that is physically shared by a small group of apartments, but is assigned for the sole use of one or few particular ones at the time, as illustrated in Figure 3 . The room is designed so as to make occupants feel they did not leave their apartment at all. They seamlessly move from their dwelling to the on-demand room and conversely, without noticing the difference, as the room adapts to their preferences.

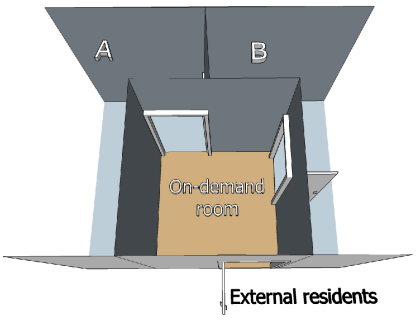


*Figure 3. Floor plan for a metamorphic house*

The underlying research problems are numerous. Dynamically "plugging" the room into a different apartment requires replacing the owner of the room's equipment, including appliances, heating, ventilation and air conditioning systems (HVAC), sensors, etc. The rights to control them and receive information from them must be dynamically reallocated. This must be done in a transparent fashion, so that off-the-shelf devices and appliances can be used.

In some cases, devices require dynamic reprogramming, like HVAC systems, because they must adapt to occupants' preferences and settings (e.g., ambient temperature set point).

Another research problem is the automatic learning of a schedule for the on-demand room. Regularities in users' requests for the room, duration of their occupation and privacy level can be discovered and learned. In this way, users do not have to manually book the room and usage conflicts can be prevented. We started investigating these research problems with an interdisciplinary approach and in collaboration with companies and public authorities [6]. We also started working on a prototype of the on-demand room solution, which will be presented as an immersive interactive virtual-reality application, leveraging the Immersia platform http://www.irisa.fr/immersia/.

## TASC Project-Team

# 6. New Results

## 6.1. Highlights of the Year

In the context of the MiniZinc Challenge and in concurrency with 16 other solvers, CHOCO has won three bronze medals in three out of four categories: free search, parallel search and Open class.

## 6.2. CHOCO

**Participants:** Jean-Guillaume Fages, Narendra Jussien, Xavier Lorca, Charles Prud'Homme.

- For second consecutive year, CHOCO has participated at the MiniZinc Challenge, an annual competition of constraint programming solvers. In concurrency with 16 other solvers, CHOCO has won three bronze medals in three out of four categories (Free search, Parallel search and Open class). Five versions have been released all year long, the last one (v3.3.0, Dec. 17th) has the particularity to be promoted on Maven Central Repository. The major modifications were related to a simplification of the API but also improvement of the overall solver.
- Within the context of the PhD thesis of Charles Prud'homme [15], a domain specific language that allows prototyping propagation engines was integrated within CHOCO, A paper appears at Constraints.
- Within the context of the PhD thesis of Charles Prud'homme [15], a generic strategy based on explanations for large neighborhood search was designed and integrated within CHOCO. A corresponding paper appears at Constraints [23].
- Within the context of the PhD thesis of Jean-Guillaume Fages, a documented package for graph variables was designed and integrated within CHOCO.

## 6.3. IBEX Solver

**Participants:** Gilles Chabert, Alexandre Goldsztejn, Bertrand Neveu, Gilles Trombettoni.

In 2014 the development on IBEX has focused on the following points:

- Rejection test based on first-order conditions (see First Order Rejection Tests For Multiple-Objective Optimization, A. Goldsztejn et al. [42]).
- Q-intersection (see Q-intersection Algorithms for Constraint-Based Robust Parameter Estimation, C. Carbonnel et al., AAAI 2014)

## 6.4. Packing curved objects

**Participants:** Nicolas Beldiceanu, Gilles Chabert, Ignacio Salas Donoso.

The development of algorithms to pack curved objects has continued. The filtering algorithm developed in 2013 for generic objects shapes has been published in the CP 2014 conference. Based on this result, we have started the design of a generic (nonlinear) packing solver in 2014. The strategy for packing is directly inspired from a successful approach recently proposed by our project partners (see On solving mixed shapes packing problems by continuous, T. Martinez et al., first BRICS countries congress on Computational Intelligence). It makes use of a stochastic optimization algorithm (CMA-ES) with a fitness function that gives a violation cost and equals zero when objects are all packed. We have generalized their approach by replacing the ad-hoc formulas (for measuring the overlapping between two objects) with an automatic calculation based on our filtering algorithm. The solver is done and the experiments have started.

## 6.5. Robustness and scheduling

**Participants:** Nicolas Beldiceanu, Mats Carlsson, Alban Derrien, Arnaud Letort, Thierry Petit, Stéphane Zampelli.

- *Robustness in the Context of the Cumulative Constraint*: This research [33] investigates cumulative scheduling in uncertain environments, using constraint programming. We get a new declarative characterization of robustness, which preserves solution quality which allow adding constraints to the main problem. In this context we adapt the 2013 sweep based algorithm in order to scale and handle several thousand of activities. We highlight the significance of our framework on a crane assignment problem with business constraints.
- *Characterization of Relevant Intervals in the Context of Energetic Reasoning*: Energetic Reasoning (ER) is a powerful filtering algorithm for the Cumulative constraint. Unfortunately, ER is generally too costly to be used in practice. One reason of its bad behavior is that many intervals are considered as relevant, although most of them should be ignored. In the literature, heuristic approaches have been developed in order to reduce the number of intervals to consider, leading to a loss of filtering. We provide a sharp characterization that allows to reduce the number of intervals by a factor seven without any loss of filtering [38].
- *Fix Point over a Conjunction of Scheduling Constraints*: This research introduces a family of synchronized sweep-based filtering algorithms for handling scheduling problems involving resource and precedence constraints. The key idea is to filter all constraints of a scheduling problem in a synchronized way in order to scale better. In addition to normal filtering mode, the algorithms can run in greedy mode, in which case they perform a greedy assignment of start and end times. The filtering mode achieves a significant speed-up over the decomposition into independent cumulative and precedence constraints, while the greedy mode can handle up to 1 million tasks with 64 resource constraints and 2 million precedences. These algorithms were implemented in both CHOCO and SICStus [21].

## 6.6. Global constraints

**Participants:** Nicolas Beldiceanu, Jean-Guillaume Fages, Xavier Lorca, Thierry Petit.

- Scalability becomes more and more critical to decision support technologies. In order to address this issue in Constraint Programming, we introduce the family of self-decomposable constraints. These constraints can be satisfied by applying their own filtering algorithms on variable subsets only. We introduce a generic framework which dynamically decompose propagation, by filtering over variable subsets. Our experiments over the cumulative constraint illustrate the practical relevance of self-decomposition [34].
- Consider a constraint on a sequence of variables functionally determining a result variable that is unchanged under reversal of the sequence. Most such constraints have a compact encoding via an automaton augmented with accumulators, but it is unknown how to maintain domain consistency efficiently for most of them. Using such an automaton for such a constraint, we derive an implied constraint between the result variables for a sequence, a prefix thereof, and the corresponding suffix. We show the usefulness of this implied constraint in constraint solving, both by local search and by propagation-based systematic search [25].
- Constraints over finite sequences of variables are ubiquitous in sequencing and timetabling. This led to general modelling techniques and generic propagators, often based on deterministic finite automata (DFA) and their extensions. We consider counter-DFAs (cDFA), which provide concise models for regular counting constraints, that is constraints over the number of times a regular-language pattern occurs in a sequence. We show how to enforce domain consistency in polynomial time for at-most and at-least regular counting constraints based on the frequent case of a cDFA with only accepting states and a single counter that can be increased by transitions. We also show that the satisfaction of exact regular counting constraints is NP-hard and that an incomplete propagator for

exact regular counting constraints is faster and provides more pruning than existing propagators. Finally, by avoiding the unrolling of the cDFA used by cost regular, the space complexity is reduced[26].

## 6.7. Optimization

**Participants:** Salvador Abreu, Alejandro Reyes Amaro, Yves Caniou, Philippe Codognet, Daniel Diaz, Jean-Guillaume Fages, Xavier Lorca, Éric Monfroy, Florian Richoux, Louis-Martin Rousseau.

- The traveling salesman problem (TSP) is a challenging optimization problem for CP and OR that has many industrial applications. Its generalization to the degree constrained minimum spanning tree problem (DCMSTP) is being intensively studied by the OR community. In particular, classical solution techniques for the TSP are being progressively generalized to the DCMSTP. Recent work on cost-based relaxations has improved CP models for the TSP. However, CP search strategies have not yet been widely investigated for these problems. The contributions of this research are twofold. We first introduce a natural generalization of the weighted cycle constraint (WCC) to the DCMSTP. We then provide an extensive empirical evaluation of various search strategies. In particular, we show that significant improvement can be achieved via our graph interpretation of the state-of-the-art Last Conflict heuristic. The work was published in the Constraints journal, see the salesman and the tree: the importance of search in CP.
- In the context of nature inspired metaheuristics and its combination with CP, some new work were conducted in the field of ant colony to solve the software project scheduling problem [19], and in the field of the Manufacturing Cell Design Problem [29].
- We implement new algorithmic methods for constraint problems on massively parallel machines. In [18], we propose an extensive study of homogeneous multi-walk parallel scheme for meta-heuristics both with and without communication. The next step will be to look at heterogeneous portfolio approaches where different solvers are looking in parallel for a solution to a given problem.

## 6.8. Modelling

**Participants:** Broderick Crawford, Frédéric Lardeux, Éric Monfroy, Ricardo Soto.

- In the framework of conversion of CST set constraints to SAT instances, a filtering engine has been studied and implemented in order to reduce the size of the generated SAT instances.
- From the one hand, CSP is very expressive. On the other hand, SAT solvers can solve huge instances (millions of variables and clauses). We thus worked on the conversion of CSP set constraints into SAT instances [35]. We then focused on the Social Golfer Problem, in order to easily integrate usual improvements (such as symmetry breaking) using our framework [40].

## 6.9. AI for real time strategy games

**Participants:** Santiago Ontanon, Florian Richoux, Alberto Uriarte.

We continue to develop an artificial intelligence, AIUR, to play the real time strategy (RTS) game $StarCraft^{tm}$, using both machine learning and constraint-based techniques. AIUR finished $4^{\text{th}}$ over 18 finalists to the $StarCraft^{tm}$ AI competition organized at the conference AIIDE 2014, and $4^{\text{th}}$ over 13 finalists to the competition at CIG 2014. This year, we wrote an ad-hoc CSPsolver to deal with the wall-in optimization problem [36] for StarCraft, and generalized it as a framework enable to handle any kind of CSP/COPmodels representing a RTS-related problem. This framework, named GHOST, helps the user to implement his CSP/COPmodel before solving it with the ready-to-use, already-tuned embedded solver.

**TEA Project-Team**

# 6. New Results

## 6.1. Highlights of the Year

This year's effort has been mainly devoted to the successful creation of project-team TEA and the definition of its new research perspective on Time, Events and Architectures in CPS design.

The SAE committee on the AADL adopted our recommendations to implement a timed and synchronous behavioural annex [13], [11] for standardisation [20]. The specification and reference implementation of this revised behavioral annex will be the focus of most our attention next year.

Adnan Bouakaz published and implemented more of the original results from his PhD. work on abstract affine scheduling [14], [15].

## 6.2. Priority-Driven Scheduling of Static Dataflow Graphs through Time Abstraction

**Participants:** Adnan Bouakaz, Thierry Gautier, Jean-Pierre Talpin.

Static dataflow graph models, such as SDF [0] and CSDF [0], are widely used to design concurrent real-time streaming applications due to their inherent functional determinism and predictable performances. The state of the art usually advocates static-periodic scheduling of dataflow graphs over dynamic scheduling. Through the past decades, a considerable effort has been made to solve this problem [0]. Ensuring boundedness and liveness is the essence of the proposed algorithms in addition to optimizing some nonfunctional performance metrics (e.g. buffer minimization, throughput maximization, etc.).

Nowadays real-time streaming applications on MPSoCs are increasingly complex; and runtime systems are more needed to handle resource sharing, task priorities, etc. Therefore, recent works [000] are considering dynamic scheduling policies (e.g. earliest-deadline first scheduling, deadline monotonic scheduling, etc.) for dataflow graphs. The main motivations of these works are: (1) most existing real-time operating systems support such scheduling policies; (2) applicability of the existing schedulability theory [00]; and (3) with such dynamic approach, multiple and independent applications, each designed as a dataflow graph, can run concurrently on the same platform.

Our work [00] [14], [15] proposes a sequence-based framework in which a large class of priority-driven schedules can be uniformly expressed and analyzed. Infinite sequences are used to describe the dataflow graphs (e.g. rate sequences, execution time sequences) and both concrete and abstract schedules (e.g. activation clocks, priority sequences, activation relations, etc.). The framework can be then easily adapted for specific needs (e.g.

[0] *Synchronous data-flow*. E. A. Lee and D. G. Messerschmitt. Proceedings of the IEEE, 1987.
[0] *Cycle-static data-flow*. Blisen, G. and Engels, M. and Lauwereins, R. and Peperstraete, Transactions on Signal Processing, v.2. 1996.
[0] *Software synthesis from dataflow graphs*. Battacharyya, S. and Lee, E. and Murthy, P. Kluwer Academic Publishers, 1996.
[0] *Affine Data-Flow Graphs for the Synthesis of Hard Real-Time Applications*. International Conference on Application of Concurrency to System Design. IEEE Press, 2012
[0] *Temporal analysis flow based on an enabling rate characterization for multi-rate applications executed on MPSoCs with non-starvation-free schedulers*. Hausmans, J., et al. International Workshop on Software and Compilers for Embedded Systems, 2014.
[0] *Hard-real-time scheduling of data-dependent tasks in embedded streaming applications.* Bamakhrama, M. and Stefanov, T. Embedded Systems Conference. ACM, 2011
[0] *Real time scheduling theory: a historical perspective*. Sha, L. et al. Real-Time Systems Conference. IEEE, 2004
[0] *A survey of hard real-time scheduling for multiprocessor systems.* Davis, R. and Burns, A. ACM Computing Surveys, v. 4, 2011
[0] *Buffer Minimization in Earliest-First Scheduling of Dataflow Graphs*. A. Bouakaz, J-P. Talpin. ACM conference on languages, compilers and tools for embedded systems. ACM Press, 2013.
[0] *Design of Safety-Critical Java Level 1 Applications Using Affine Abstract Clocks*. A. Bouakaz, J-P. Talpin. International Workshop on Software and Compilers for Embedded Systems, 2013.

affine scheduling). Our schedule construction approach is based on two steps. The first step consists in computing an abstract schedule which consists of a set of priority sequences, processor allocation sequences, and activation relations. An activation relation between two actors describes the relative order of their activations, and hence allows us to compute safe sizes of channels between them using worst-case overflow/underflow scenarios. This step must satisfy some correctness constraints such as consistency and exclusion of overflow and underflow exceptions. Once the best abstract schedule (w.r.t. to a performance metric) is computed, the schedule is refined by computing the actual periods and phases that ensure schedulability on the target architecture.

## 6.3. Formal Verification of a Synchronous Data-flow Compiler: from Signal to C

**Participants:** Van-Chan Ngo, Jean-Pierre Talpin, Thierry Gautier, Paul Le Guernic, Loïc Besnard.

Translation validation [00] is a technique that attempts to verify that program transformations preserve the program semantics. It is obvious to prove globally that the source program and its final compiled program have the same semantics. However, we believe that a better approach is to separate concerns and prove each analysis and transformation stage separately with respect to ad-hoc data-structures to carry the semantic information relevant to that phase.

In the case of the Signal compiler [1], [7][12], the preservation of the semantics can be decomposed into the preservation of clock semantics at the *clock calculation* phase and that of data dependencies at the *static scheduling* phase, and, finally, value-equivalence of variables at the *code generation* phase.

**Translation Validation for Clock Transformations in a Synchronous Compiler.** In this work, the clock semantics of the source and transformed programs are formally represented as *clock models*. A clock model is a first-order logic formula that characterizes the presence/absence status of all signals in a Signal program at a given instant. Given two clock models, a *clock refinement* between them is defined which expresses the semantic preservation of clock semantics. A method to check the existence of clock refinement is defined as a satisfiability problem which can be automatically and efficiently proved by a SMT solver.

Let $Cp^{sig}$ and $Val_{clk}$ be the functions which define the Signal compiler and a validator, respectively. The following function defines a formally verified compiler for the *clock calculation and Boolean abstraction* phase. We write $C \sqsubseteq_{clk} A$ to denote that there exists a refinement between $A$ and $C$.

$$Cp^{sig}_{Val_{clk}}(A) = \begin{cases} C & \text{if } Cp^{sig}(A) = C \text{ and } Val_{clk}(A, C) = true \\ Error & \text{if } Cp^{sig}(A) = C \text{ and } Val_{clk}(A, C) = false \\ Error & \text{if } Cp^{sig}(A) = Error \end{cases}$$

where $Val_{clk}(A, C) =$ true if and only if $C \sqsubseteq_{clk} A$.

**Precise Deadlock Detection for Polychronous Data-flow Specifications.** Dependency graphs are a commonly used data structure to encode the streams of values in data-flow programs and play a central role in scheduling instructions during auto-mated code generation from such specifications. In this work [17], we propose a precise and effective method that combines a structure of dependency graph and first order logic formulas to check whether multi-clocked data-flow specifications are deadlock free before generating code from them. We represent the flow of values in the source programs by means of a dependency graph and attach first-order logic formulas to condition these dependencies. We use an SMT solver [0] to effectively reason about the implied formulas and check deadlock freedom.

[0] *Translation validation.* Pnueli A., Siegel M., and Singerman E. In Proceedings of TACAS'98, 1998.
[0] *Translation validation: From signal to c.* M. Siegel A. Pnueli and E. Singeman. In Correct Sytem Design Recent Insights and Advances, 2000.
[0] *Satisfiability modulo theories: An appetizer.* L. de Moura and N. Bjorner. In Brazilian Symposium on Formal Methods, 2009.

**Evaluating SDVG translation validation: from Signal to C.** This work focuses on proving that every output signal in the source program and the corresponding variable in the compiled program, the generated C program, have the same values. The computations of all signals and their compiled counterparts are represented by a shared value-graph, called *Synchronous Data-flow Value-Graph* (SDVG).

Given a SDVG, assume that we want to show that two variables have the same value. We simply need to check that they are represented by the same sub-graph, meaning that they point to the same graph node. If all output signals in the source program $A$ and the corresponding variables in the generated $C$ program have the same value, then we say that $C$ refines $A$, denoted by $C \sqsubseteq_{val} A$.

**Implementation and Experiments**. At a high level, our tool *SigCert* (https://scm.gforge.inria.fr/svn/sigcert) developed in OCaml checks the correctness of the compilation of Signal compiler w.r.t clock semantics, data dependence, and value-equivalence as given in Figure 3 .

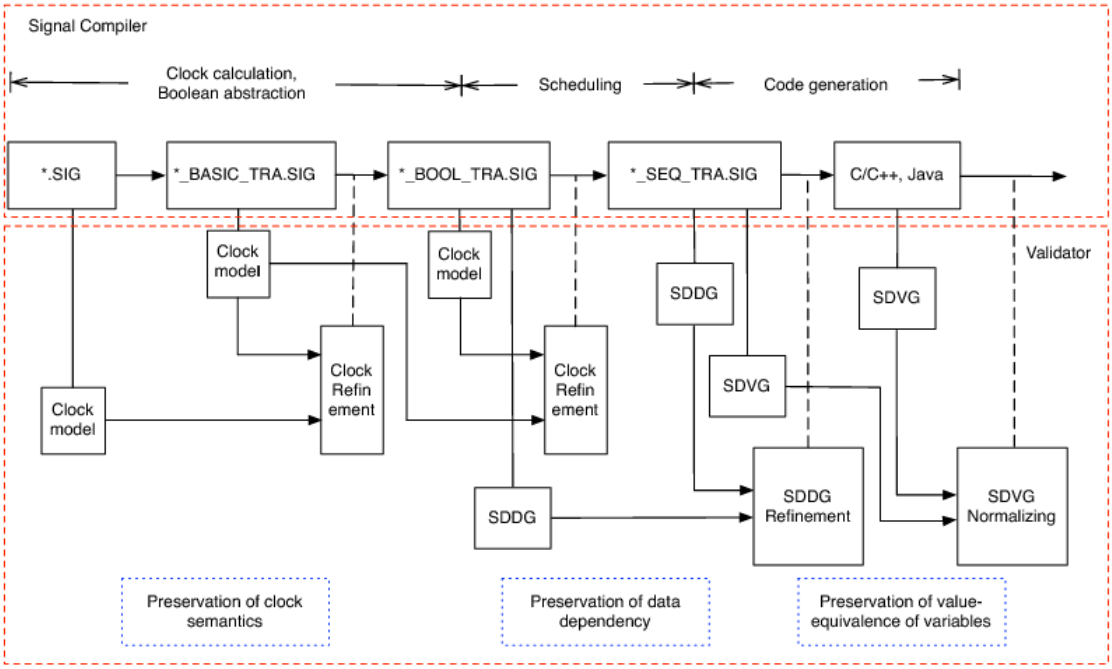


*Figure 3. Our Integration within Polychrony Toolset*

## 6.4. Ongoing integration of Polychrony with the P toolset

**Participants:** Christophe Junke, Loïc Besnard, Thierry Gautier, Paul Le Guernic, Jean-Pierre Talpin.

**Current state of P.** The FUI project P has been extended until September 2015. Partners in the project now focus on code generation aspects, leaving software architecture aspects aside. The qualifiable model-based code generator, previously known as P toolset, is now named QGen (QGen is developped mostly in Ada 2012 and Python).

**Model transformation (P2S).** We developped a transformation tool hereafter named `P2S` for expressing P system models as Signal processes. Our work is based on EMF (Eclipse Modelling Framework), taking advantage of the existing Ecore metamodels available for both P and SSME.

The P2S tool is written in Clojure, which is a dialect of Lisp running on the Java Virtual Machine. This approach allows to benefits from a terse and expressive language while remaining fully interoperable with existing Java libraries (including Eclipse plugins and especially Polychrony ones).

**SSME abstraction layer.** P2S uses an abstraction layer to simplify the creation of SSME elements, while taking into account EMF idioms. For example, the following expression creates a `ProcessModel` instance using the currently registered EMF factory:

```
(process "TestProcess"
         :in '[boolean h integer x]
         :out '[integer y]
         :body (sigdef
                (id 'y)
                (when* (id 'x) (id 'h))))
```

The newly created object can be saved as an XMI file using EMF utilities (the XMI file is 40 lines long and not shown here). This object and its children represent the following Signal process expression [0]:

```
process TestProcess =
  (? boolean h; integer x;
   ! integer y; )
  (| y := (x when h) |);
```

**Transformation to P.** Conversion from P to Signal relies on Clojure's multimethods. We defined a `convert` multimethod which dispatches on the type of its argument and possibly on additional modifiers. This mechanism allows to convert expressions differently depending on whether we want to produce a Signal declaration or an expression. For example, the following method specializer converts a P port as a signal declaration:

```
(defmethod convert [Port :declaration] [port & _]
  (ssme/signal-declarations
   (convert (.getDataType port))
     (ssme/with-comment
       [(readable-name port :declaration) :post]
       (ssme/id (p-name port)))))
```

Since the specializer contains the `:declaration` keyword, the previous conversion is applied only when called with that keyword given as an extra argument, as follows:

```
(convert some-port :declaration)
```

The more general specializer, which is defined below, is meant to be used inside Signal expressions and, as such, only returns a Signal identifier:

```
(defmethod convert Port [port]
   (ssme/id (p-name port)))
```

Note also that thanks to class inheritance, the above methods are sufficient to convert all kind of P ports (input/output, data/control).

The naming scheme for the resulting SSME elements is handled by the `p-name` multi-method and relies on XMI identifiers of the original P elements: XMI identifiers generated by QGen are string representations of positive integers. Moreover, those identifiers are guaranteed to be unique in a model. These two properties allows to generate valid Signal identifiers while ensuring traceability (e.g. signal `P101` links to the unique port of the original model having 101 as a unique identifier).

Datatypes are currently converted as Signal predefineds types, which do not always match exactly the original types. Another partially implemented option consists in translating them as `external` types in Signal. Some types, like arrays, are converted the same way with both approaches:

[0] Even using the dedicated `signalTreeAPI` utility class, the same example would require many more lines of Java code.

```
(defmethod convert TArray [a]
  (reduce (fn [base dim]
            (ssme/array-type base (convert dim :signal)))
          (convert (.getBaseType a))
          (.getDimensions a)))
```

Conversion of arithmetic operations may also lead to predefined Signal operators (by default) or externally defined functions (incomplete). The current approach has been tested on QGen's test models and successfully translates 208 of the 227 models.

**Partial block sequencing.** The conversion from P models to Signal takes into account block dependencies as computed by QGen. Unfortunately, QGen's block sequencer produces a total order between blocks, with leads to over-constrained Signal models. We contributed to the model compiler by writing an alternative (Ada) package which provides: (i) a way to parameterize block sequencing, and (ii) partial ordering options.

Our implementation is not part of the qualified compiler, but available as a standalone (non-qualifiable) executable. However, during the development of this block sequencer, we were able to find and correct existing bugs in QGen's sequencer.

**Perspectives.** From a software development point of view, our current work needs to be packaged and better integrated with the build system of Polychrony. By the way, that existing build process itself could be slightly improved by using Maven configuration files instead of Eclipse manual plug-in management.

The use of a functional language on top of the Java Virtual Machine is an interesting aspect of our work. By allowing the abstraction layer, which currently works at the SSME level, to also access the existing Signal library, we could provide an API for writing and compiling Signal code using a domain-specific language expressed in Clojure (there already exist JNI bindings with the native library). This feature could help developpers hook into, or interact with, the existing Signal compiler in order to customize parts of the code generation strategies.

Regarding the P project, we still need to test code distribution strategies on industrial use-cases and determine how it can be exploited at the system-model level.

## 6.5. A synchronous annex for the AADL

**Participants:** Loïc Besnard, Thierry Gautier, Paul Le Guernic, Jean-Pierre Talpin.

The SAE committee on the AADL adopted our recommendations to implement a timed and synchronous behavioural annex for the standard [20]. The specification and reference implementation of this revised behavioral annex will be the focus of most our attention next year.

We propose a synchronous timing annex for the SAE standard AADL. Our approach consists of building a synchronous model of computation and communication that best fits the semantics and expressive capability of the AADL and its behavioral annex and yet requires little to know (syntactic) extension to it, i.e. to identify a synchronous core of the AADL (which prerequisites a formal definition of synchrony at hand) and define a formal design methodology to use the AADL in a way that supports formal analysis, verification and synthesis.

Our approach first identifies the core AADL concepts from which time events can be described. Then, is considers the behavior annex (BA) as the mean to model synchronous signals and traces through automata. Finally, we consider elements of the constraint annex to reason about abstractions of these signals and traces by clocks and relations among them. To support the formal presentation of these elements, we define a model of automata that comprises a transition system to express explicit transitions and constraints, in the form of a boolean formula on time, to implicitly constraint its behavior. The implementation of such an automaton amounts to composing its explicit transition system with that of the controller synthesised from its specified constraints.

## 6.6. New features of Polychrony

**Participants:** Loïc Besnard, Thierry Gautier, Paul Le Guernic.

**Reduction of communications.** We have developed, as a general functionality of the Signal toolbox, a means to reduce communications between two graphs, using assignment clocks and utility clocks.

For a given signal $x$, its assignment clock represents the instants at which it may be modified (otherwise than keeping its previous value $x\$$) while its utility clock in a given graph represents the instants at which it is effectively used in this graph.

Considering two graphs $G_i$ and $G_j$ with a signal $x$ sent from $G_i$ to $G_j$, containers are built above $G_i$ and $G_j$ in order to minimize the clock at which $x$ must be communicated. On the sender side, the signal which has to be sent can be reduced to $x_j$ with $x_j := x \texttt{ when } h$, where $h$ is the lower bound of the assignment clock of $x$ and the utility clock of $x$ in $G_j$. On the receiver side, $x$ is replaced in $G_j$ by $x\_r$ with $x\_r := x_j \texttt{ default } x\_r\$$.

Note that this reduction is not always possible because it may introduce cycles between signals and clocks.

Experiments have been made on programs intended to the distribution of Quartz applications, with a gain of up to 40on some of them [18].

**Polychronous automata.** We have defined a new model of polychronous constrained automata that has been provided as semantic model for our proposal of an extension of the AADL behavioural annex [20]. An algebra of regular expressions is also defined to represent abstractions of constrained automata or, more specifically, their time constraints.

An experimental implementation of the semantic features of this "timing annex" will be provided through the Polychrony framework. For that purpose, representations of automata are introduced in the Signal toolbox of Polychrony. In a first step, we have decided to provide only a minimal extension of the Signal language itself. A new syntactic category of process model, which is an automaton model, has been introduced. States are described by the association of labels with subprocesses, as it is available in Signal, and transitions between states, at a given clock, are written as calls to *intrinsic* (predefined) processes. Constraints described as regular expressions on events should also be introduced using intrinsic processes.

Automata will be used in different ways related to stategies of compilation. In particular, they will serve as an alternative model for the code generation. For that purpose, polychronous programs are rewritten thanks to valuations of memorized boolean signals. The resulting partially valuated programs are the states of a control automaton.

Such techniques can be applied to implement endo-isochronous programs. Currently, code may be generated only for endochronous programs, for which clock hierarchy is a tree. Endo-isochronous programs are compositions of endochronous programs the "intersection" of which is also endochronous. For example, an automaton can be built to generate code when two signals are known to alternate.

## 6.7. Optimized Distribution of Synchronous Programs via a Polychronous Model

**Participants:** Ke Sun, Jean-Pierre Talpin, Thierry Gautier, Loïc Besnard.

We propose a distribution methodology for synchronous programs [18], applied in particular on programs written in the Quartz language [0]. The given program is first transformed into an intermediate model of guarded actions. After user-specified partitioning, the generated sub-models are transformed into equivalent Signal processes [7]. Then, the unnecessary constraints are eliminated from the processes to avoid unnecessary synchronization. Finally, within the Signal framework, the minimal frequencies of communication and computation are computed via multi-clock calculation. This operation can efficiently reduce the communication quantity and the computation load, with no change to the interface behaviors. Along this way, an optimized data-flow network over desynchronized processing locations can be constructed.

[0] *The Synchronous Programming Language Quartz*. K. Schneider, Technical Report n. 375. University of Kaiserslautern, 2009

The presented methodology has been implemented within the integrated framework Quartz/Averest + Signal/Polychrony. To illustrate and validate this methodology, a series of examples served as case studies. Each of them has been written in the Quartz language and distributed over different processing locations using the presented optimization methodology. These case studies confirm that the optimization can bring in significant communication reduction. In the sequel, the efficient utilization of distributed systems is substantially updated.

## 6.8. Component-based Design of Multi-rate Systems

**Participants:** Ke Sun, Jean-Pierre Talpin, Thierry Gautier, Loïc Besnard.

The Synchronous language Quartz is well suited for modeling mono-clocked systems. However, as based on the model of computation (MoC) synchrony, its parallelism feature excessively strengthens the synchronization. Such synchronous parallelism in particular restricts independent component design. That is, the modeling of connected components should constantly refer to each other to guarantee the achievement of desired system behavior. Hence, Quartz cannot support well the component-based system design, in particular for the distributed systems that are generally deployed over desynchronized processing locations with multi-rate clocks.

In contrast to Quartz, the polychronous language Signal is based on the MoC polychrony. As its name suggests, a polychronous program makes use of multi-rate clocks to drive its execution. One can consider that each component in the program holds its own master clock, and there is no longer a master clock for the whole program. The resulted architecture is named globally asynchronous locally synchronous (GALS) architecture.

Through integrating Quartz with Signal, a component-based methodology is proposed for designing multi-rate systems: at first, components are modeled independently to achieve local behaviors; secondly, inter-component communications are adjusted using Signal to realize intermittent synchronization. In this way, the modeling approach for mono-clocked systems evolves into a component-based modeling methodology. Such significant progress not only facilitates the component coordination, but also enhances the component reusability, in particular for modeling large scale systems.

**VISAGES Project-Team**

# 6. New Results

## 6.1. Highlights of the Year

Dr Camille Maumet was awarded by the French Society of Magnetic Resonance in Biology and Medicine (SFRMBM) for her PhD Thesis on analysis of neuroimaging data including images from functional Magnetic Resonance Imaging (fMRI) and Arterial Spin Labeling http://www2.warwick.ac.uk/fac/sci/wmg/idh/idhnews/?tag=Neural+Engineering.

Dr Americ Stamm was awarded by the Univ. of Rennes I foundation as the best PhD thesis in Math, Computer Sciences and Electrical Engineering. This award is dedicated for the PhDs having the highest potential for innovation and technological transfer
;ses-de-la-fondationhttps://fondation.univ-rennes1.fr/les-prix-de-thèses-de-la-fondation.

## 6.2. Image Computing: Detection, Segmentation, Registration and Analysis

### *6.2.1. A Mathematical Framework for the Registration and Analysis of Multi-Fascicle Models for Population Studies of the Brain Microstructure*

**Participant:** Olivier Commowick.

Diffusion tensor imaging (DTI) is unable to represent the diffusion signal arising from multiple crossing fascicles and freely diffusing water molecules. Generative models of the diffusion signal, such as multi-fascicle models, overcome this limitation by providing a parametric representation for the signal contribution of each population of water molecules. These models are of great interest in population studies to characterize and compare the brain microstructural properties. Central to population studies is the construction of an atlas and the registration of all subjects to it. However, the appropriate definition of registration and atlasing methods for multi-fascicle models have proven challenging. This paper proposes [24] a mathematical framework to register and analyze multi-fascicle models. Specifically, we define novel operators to achieve interpolation, smoothing and averaging of multi-fascicle models. We also define a novel similarity metric to spatially align multi-fascicle models. Our framework enables simultaneous comparisons of different microstructural properties that are confounded in conventional DTI. The framework is validated on multi-fascicle models from 24 healthy subjects and 38 patients with tuberous sclerosis complex, 10 of whom have autism. We demonstrate the use of the multi-fascicle models registration and analysis framework in a population study of autism spectrum disorder. This work was performed in close collaboration with the Children's Hospital Boston.

### *6.2.2. Longitudinal Intensity Normalization in Multiple Sclerosis Patients*

**Participants:** Yogesh Karpate, Olivier Commowick, Christian Barillot, Gilles Edan.

In recent years, there have been many Multiple Sclerosis studies using longitudinal MR images to study and characterize the MS lesion patterns. The intensity of similar anatomical tissues in MR images is often different because of the variability of the acquisition process and different scanners. We proposed [29] a novel methodology for a longitudinal lesion analysis based on intensity standardization to minimize the inter-scan intensity difference. The intensity normalization maps parameters obtained using a robust Gaussian Mixture Model (GMM) estimation not affected by the presence of MS lesions. Experimental results demonstrated that our technique accurately performs the task of intensity standardization. We show consequently how the same technique can improve the results of longitudinal MS lesion detection.

### *6.2.3. Simultaneous Estimation of T1, T2 and B1 Maps From Relaxometry MR Sequences*

**Participants:** Fang Cao, Olivier Commowick, Elise Bannier, Christian Barillot.

Interest in quantitative MRI and relaxometry imaging is rapidly increasing to enable the discovery of new MRI disease imaging biomarkers. While DESPOT1 is a robust method for rapid whole-brain voxelwise mapping of the longitudinal relaxation time (T1), the approach is inherently sensitive to inaccuracies in the transmitted flip angles, defined by the B1 inhomogeneity field, which become more severe at high field strengths (e.g., 3T). We have proposed [26] a new approach for simultaneously mapping the B1 field, M0 (proton density), T1 and T2 relaxation times based on regular fast T1 and T2 relaxometry sequences. The new method is based on the intrinsic correlation between the T1 and T2 relaxometry sequences to jointly estimate all maps. It requires no additional sequence for the B1 correction. We evaluated our proposed algorithm on simulated and in-vivo data at 3T, demonstrating its improved accuracy with respect to regular separate estimation methods.

### *6.2.4. Quantitative Relaxation Templates for the Human Brain at 3T*

**Participants:** Fang Cao, Olivier Commowick, Christian Barillot.

Quantitative MRI (qMRI) templates of relaxation times and proton density can be of particular interest for dedicated clinical applications such as characterizing brain tissue abnormalities, as well as general research purposes. We have developed in [27] 3D qMRI statistical templates consisting of T1, T2, T2* and $\rho$* maps from the human brain at 3T. The qMRI templates were built from a population of 20 normal controls, for which individual maps were estimated in a robust manner, accounting for acquisition artifacts and expected relationships between the relaxometry parameters. For validation, we fed the qMRI templates into a realistic MRI simulator to synthesize MR-weighted images, and compared these images with the real MR acquisitions. High correlation coefficients (>0.80) show that the developed qMRI templates can be used as input dataset for MRI simulation community, which may be of great interest to clinical neuroscience field.

### *6.2.5. Myelin Water Fraction Imaging in Multiple Sclerosis patients*

**Participants:** Olivier Commowick, Elise Bannier, Christian Barillot.

Multi-echo T2 relaxometry is a relevant imaging method for Myelin Water Fraction (MWF) quantification in the study of multiple sclerosis (MS). However, to ensure accurate estimation, a large number of echoes are still required that can drive to very long acquisitions. In practice, 32 echo times (TE) ranging from 10 ms to 320 ms and an echo spacing (ESP) of 10 ms are used. Analysis of the decay curve of the consecutive echoes allows the estimation of the T2 spectrum. The proposed approach makes use of recent spatial regularization methods for MWF estimation from clinically compatible acquisitions (typically 11 echoes acquired within 6 minutes with TE1=ESP=8.4 ms). The algorithms were evaluated on both synthetic and clinical data, illustrating the ability to compute accurate MWF maps from a low number of echoes. The 2 methods used a priori information as well as conventional and fast algorithm (NNLS), and a cross-validation strategy. Based on simulated and clinical data results, the nlsrNNLS estimation is more accurate and less penalizing than srNNLS. This regularization provides an efficient way to circumvent an ill-posed problem aspect, in particular with a reduced number of echoes for clinically acceptable acquisition times, allowing for accurate MWF estimation. This work, performed in the master internship of Lucas Soustelle, was accepted as a conference abstract at SFRMBM 2015, and is submitted to ISMRM.

## 6.3. Image processing on Diffusion Weighted Magnetic Resonance Imaging

### *6.3.1. Fast Identification of Optimal Fascicle Configurations from Standard Clinical Diffusion MRI Using Akaike Information Criterion*

**Participants:** Olivier Commowick, Christian Barillot.

Analytic multi-compartment models have gained a tremendous popularity in the recent literature for studying the brain white matter microstructure from diffusion MRI. This class of models require the number of compartments to be known in advance. In the white matter however, several non-collinear bundles of axons, termed fascicles, often coexist in a same voxel. Determining the optimal fascicle configuration is a model selection problem. We have proposed [30], [33] a novel approach to identify such a configuration from clinical diffusion MRI where only few diffusion images can be acquired and time is of the essence. Starting from a

set of fitted models with increasing number of fascicles, we use Akaike information criterion to estimate the probability of each candidate model to be the best Kullback-Leibler model. These probabilities are then used to average the different candidate models and output an MCM with optimal fascicle configuration. This strategy is fast and can be adapted to any multi- compartment model. We illustrate its implementation with the ball-and-stick model and show that we obtain better results on single-shell low angular resolution diffusion MRI, compared to the state-of-the-art automatic relevance detection method, in a shorter processing time.

### 6.3.2. Tracking the Cortico-Spinal Tract as a Multi-Modal Distribution of Streamlines from Local White Matter Microstructure Models

**Participant:** Olivier Commowick.

We have presented [31] a pipeline to reconstruct the corticospinal tract (CST) that connects the spinal cord to the motor cortex. The proposed method combines a new white matter microstructure model coined Diffusion Directions Imaging and a new tractography algorithm based on a particle filter adapted for approximating multi-modal distributions. In this paper, we put the computation time and accuracy of our pipeline to the test in the context of the MICCAI 2014 DTI challenge, which aims to provide fast and accurate reconstructions of the CST for presurgical planning of brain tumor extraction. These two key performance metrics are expected in such a situation where time is of the essence and the quality of the data is dependent on the patient's health condition and ability to cooperate. In no more than 1.5 hours per patient, we successfully provide accurate CSTs of 2 very collaborative patients who underwent a diffusion MRI protocol that included 69 diffusion-sensitizing gradients spread over 4 different shells ranging from b = 200 to b = 3000 $s/mm^2$.

### 6.3.3. Model selection improvement with non-central chi estimation of multi-compartment models

**Participant:** Olivier Commowick.

Diffusion images are known to be corrupted with a non-central chi (NCC)-distributed noise. There has been a number of proposed image denoising methods that account for this particular noise distribution. However, to the best of our knowledge, no study was performed to assess the influence of the noise model in the context of diffusion model estimation. In particular, multi-compartment models are an appealing class of models to describe the white matter microstructure but require the optimal number of compartments to be known a priori. Its estimation is no easy task since more complex models will always better fit the data, which is known as over-fitting. However, MCM estimation in the literature is performed assuming a Gaussian-distributed noise. We have shown in a preliminary study [32] that using the appropriate NCC distribution for modeling the noise model reduces significantly the over-fitting, which could be helpful for unraveling model selection issues and obtaining better model parameter estimates.

### 6.3.4. Symmetric Block-Matching Registration for the Distortion Correction of Echo-Planar Images

**Participants:** Renaud Hédouin, Olivier Commowick, Elise Bannier, Christian Barillot.

We have introduced a new approach to correct geometric and intensity distortion of Echo Planar Images (EPI) from images acquired with opposite phase encoding directions. A new symmetric block-matching registration algorithm has been developed for this purpose relying on new adapted transformations between blocks and a symmetric optimization scheme to ensure an opposite symmetric transformation. We present results of our algorithm showing its ability to robustly recover EPI distortion while obtaining sharper results than the popular TOPUP algorithm. This work was performed in close collaboration with the Children's hospital in Boston.

## 6.4. Medical Image Computing in Brain Pathologies

### 6.4.1. Adaptive Dictionary Learning for Competitive Classification of Multiple Sclerosis Lesions

**Participants:** Hrishikesh Deshpande, Pierre Maurel, Christian Barillot.

The manual delineation of Multiple Sclerosis (MS) lesions is a challenging task pertaining to the requirement of neurological experts and high intra- and inter-observer variability. It is also time consuming because large number of Magnetic Resonance (MR) image slices are needed to obtain 3-D information. Over the last years, various models combined with supervised or unsupervised classification methods have been proposed for segmentation of MS lesions using MR images. Recently, signal modeling using sparse representations (SR) has gained tremendous attention and is an area of active research. SR allows coding data as sparse linear combinations of the elements of over-complete dictionary and has led to interesting image recognition results. In this work, we have proposed to use a sparse representation and an adaptive dictionary learning paradigm to automatically classify Multiple Sclerosis (MS) lesions from MRI. In particular, we investigate the effects of learning dictionaries specific to the lesions and individual healthy brain tissues, which include White Matter (WM), Gray Matter (GM) and Cerebrospinal Fluid (CSF). The dictionary size plays a major role in data representation but it is an even more crucial element in the case of competitive classification. We present an approach that adapts the size of the dictionary for each class, depending on the complexity of the underlying data. The proposed algorithm is evaluated on 3-D multi-channel MR images demonstrating improved classification.

### *6.4.2. Predictive Value of Imaging Markers at Multiple Sclerosis Disease Onset Based on Gadolinium- and USPIO- Enhanced MRI and Machine Learning*

**Participants:** Olivier Commowick, Jean-Christophe Ferré, Elise Bannier, Gilles Edan, Christian Barillot.

A novel characterization of Clinically Isolated Syndrome (CIS) patients according to lesion patterns has been proposed in [13]. More specifically, patients are classified according to the nature of inflammatory lesions patterns. It is expected that this characterization can infer new prospective figures from the earliest imaging signs of Multiple Sclerosis (MS), since it can provide a classification of different types of lesions across patients. The method is based on a two-tiered classification. Initially, the spatio-temporal lesion patterns are classified. The discovered lesion patterns are then used to characterize groups of patients. The patient groups are validated using statistical measures and by correlations at 24-month follow-up with hypointense lesion loads. The methodology identified 3 statistically significantly different clusters of lesion patterns showing p-values smaller than 0.01. Moreover, these patterns defined at baseline correlated with chronic hypointense lesion volumes by follow-up with an $R^2$ score of 0.90. The proposed methodology is capable of identifying three major different lesion patterns that are heterogeneously present in patients, allowing a patient classification using only two MRI scans. This finding may lead to more accurate prognosis and thus to more suitable treatments at early stage of MS.

### *6.4.3. Robust detection of multiple sclerosis lesion from intensity-normalized multi-channel MRI*

**Participants:** Yogesh Karpate, Olivier Commowick, Christian Barillot.

Multiple sclerosis (MS) is a disease with heterogeneous evolution among the patients. Better understanding of the disease will lead to improved patient-adapted therapeutic strategies. We propose a novel paradigm to detect MS lesions based on a statistical framework which consists of detection based on differences between multi-channel MRI of patients and controls. This framework fused with intensity standardization was applied to the study of MS and highlighted the great interest of quantitative MRI measurements for a better characterization of MS. Experimental results demonstrate that our technique accurately detects significant differences in lesions consequently improving the results of MS lesion detection. This work has been accepted to SPIE Medical Imaging 2015.

### *6.4.4. Multiple Sclerosis Lesions Recognition: One Class Learning Approach*

**Participants:** Yogesh Karpate, Olivier Commowick, Christian Barillot, Gilles Edan.

We have developed an automatic algorithm for the detection of multiple sclerosis lesions (MSL) from multi-sequence magnetic resonance imaging (MRI). We build a probabilistic classifier that can recognize MSL as a novel class, trained only on Normal Appearing Brain Tissues (NABT). Patch based intensity information of MRI images is used to train a classifier at the voxel level. The classifier is in turn used to compute a probability characterizing the likelihood of each voxel to be a lesion. This probability is then used to identify a lesion voxel based on simple Otsu thresholding. This work has been submitted to ISBI 2015.

## 6.5. Vascular Imaging and Arterial Spin Labeling

### *6.5.1. Peripheral angiography and neurovascular imaging*

**Participants:** Hélène Raoult, Jean-Yves Gauvrit, Elise Bannier, Pierre Maurel, Christian Barillot, Jean-Christophe Ferré.

Work-in-progress Non contrast enhanced MR angiography sequences were optimized on phantom as well as healthy volunteers and evaluated on patients presenting arterio venous malformations (AVM). High temporal resolution (70ms) images were obtained and compared to the gold standard Digital Subtraction Angiography. Results showed that Time-resolved SL MR angiographic imaging over two cardiac cycles is a reliable clinical tool for cerebral AVM characterization, yielding very good to excellent agreement with DSA. This work was published in Radioloy late 2013. This data was also post processed to obtain hemodynamics maps (time to peak, wash-in, wash out and mean transit time) and discriminate among different AVM components to relate hemodynamic patterns with rupture risk. This work was published in Stroke [23].

### *6.5.2. Robust perfusion maps in Arterial Spin Labeling by means of M-estimators*

**Participants:** Pierre Maurel, Jean-Christophe Ferré, Christian Barillot.

Non-invasive measurement of Cerebral Blood Flow (CBF) is now feasible thanks to the introduction of Arterial Spin Labeling (ASL) Magnetic Resonance Imaging (MRI) techniques. To date, due to the low signal-to-noise ratio of ASL, a single acquisition (pair of control/label scans) is not sufficient to estimate perfusion reliably. Instead, the acquisition is usually repeated several times and the perfusion information is calculated by averaging across the repetitions. However, due to its zero breakdown point, the sample mean is very sensitive to outliers. We have proposed [18] to compute ASL CBF maps using Huber's M-estimator, a robust statistical function that is not overly impacted by outlier. This work was part of the PhD thesis of Camille Maumet.

### *6.5.3. Brain perfusion gender difference study using MRI in young adults*

**Participants:** Léa Itmi, Pierre Maurel, Christian Barillot.

The usage of population models is becoming increasingly important in cerebral imaging, particularly in ASL. Therefore, it is important to check the limits of the models before applying them, to guarantee the reliability of the results. It is now well-known that brain perfusion changes with the age, and this effect is taken into account when evaluating brain perfusion images. But gender differences have not been well studied yet. It is known for a long time that female brain perfusion is higher than male brain perfusion, but few studies have investigated whether some regional perfusion differences exist or not. We evaluate whether, as for the age, gender differences should be taken into account when analyzing brain perfusion images. We focus on young adults subjects and studied, at the region level and the voxel level if gender differences exist and how it differs. The overall and regional differences were analyzed and then we also investigated the perfusion asymmetries in the brain (left hemisphere versus right hemisphere).

## 6.6. EEG and MR Imaging

### *6.6.1. Feasibility and specificity of simultaneous EEG and ASL MRI at 3T*

**Participants:** Elise Bannier, Marsel Mano, Isabelle Corouge, Lorraine Perronnet, Christian Barillot.

Brain functional imaging can be performed using several approaches, including EEG, BOLD and ASL MRI. The Neurinfo platform has acquired an EEG MR compatible 64ch device over the summer to perform joint EEG and BOLD or ASL fMRI. To date, only a few studies have addressed the issue of connecting EEG signal to ASL perfusion. The aim of this study was to assess ASL-EEG at 3T in terms of safety as well as EEG and MR signal quality. The temperature measurements, specific absorption rate, and signal to noise ratio experiments have shown that ASL EEG can be safely performed using the parameters presented above. However, residual gradient artifacts in the PASL-EEG data have to be considered. Further research is needed to understand the artifact variability and to develop an appropriate correction strategy. This study is performed as part of the HEMISFER project in close collaboration with the involved teams.

### *6.6.2. Neurofeedback using Virtual Reality and Hybrid EEG-MRI for Brain Rehabilitation*

**Participants:** Lorraine Perronnet, Marsel Mano, Christian Barillot.

We have conducted a thorough state-of-the-art of Neurofeedback (NF) and restorative Brain Computer Interfaces (BCI) under EEG and fMRI modality as well as of EEG-fMRI integration, with a particular focus on applications in depression and motor rehabilitation. This enabled us to build a theoretical comparison of EEG- and fMRI-NF methodology that will be helpful in designing NF protocols combining both modalities. In this perspective, we are currently designing a NF protocol based on motor imagery that will be compatible with EEG and fMRI, and running preliminary recordings of motor execution and motor imagery. Besides, we are writing a book chapter about NF and BCI that is intended to disambiguate the existing definitions and to present basic knowledge about NF principles and applications to naive readers. This is a joint work with Anatole Lécuyer team (Hybrid), in the frame of the Hemisfer project.

### *6.6.3. Symmetrical EEG and fMRI Imaging by Sparse Regularization*

**Participants:** Thomas Oberlin, Pierre Maurel, Christian Barillot.

This work considers the problem of brain imaging using simultaneously recorded electroencephalography (EEG) and functional magnetic resonance imaging (fMRI). To this end, we introduce a linear coupling model that links the electrical EEG signal to the hemodynamic response from the blood-oxygen level dependent (BOLD) signal. Both modalities are then symmetrically integrated, to achieve a high resolution in time and space while allowing some robustness against potential decoupling of the BOLD effect. The novelty of the approach consists in expressing the joint imaging problem as a linear inverse problem, which is addressed using sparse regularization. The sparsity prior naturally reflects the fact that only few areas of the brain are activated at a certain time, and it is easily implemented through efficient so-called proximal algorithms. The significance of the method and the effectiveness of the algorithms are demonstrated through numerical investigations on a simplified head model and simulated data on a realistic brain model. This is a joint work with Remi Gribonval team (Panama), in the frame of the Hemisfer project.